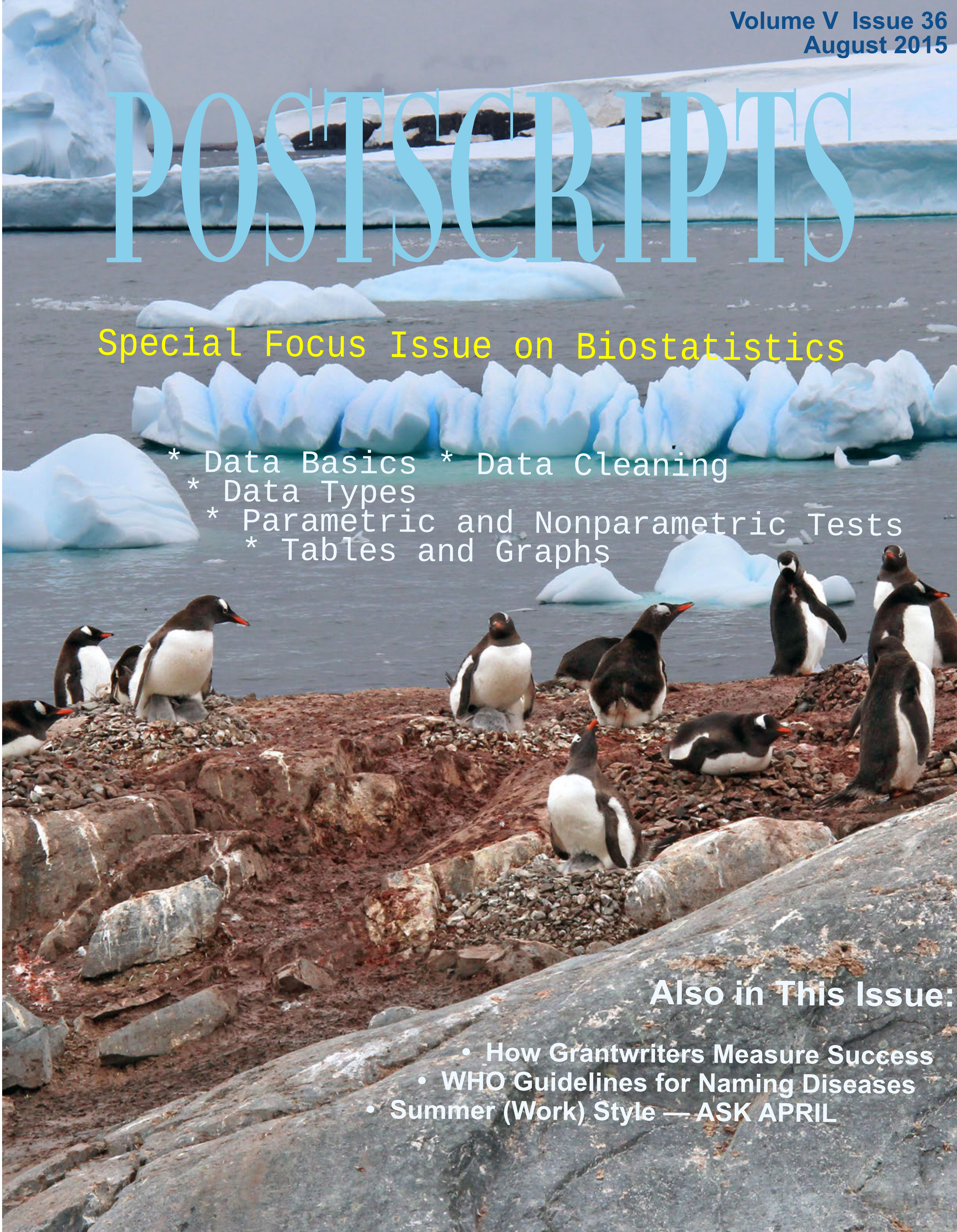
Volume V Issue 36
August 2015

POSTSCRIPTS



Special Focus Issue on Biostatistics

* Data Basics * Data Cleaning
* Data Types
* Parametric and Nonparametric Tests
* Tables and Graphs

Also in This Issue:

- How Grantwriters Measure Success
- WHO Guidelines for Naming Diseases
- Summer (Work) Style — ASK APRIL

Postscripts August 2015 AMWA Pacific Southwest Chapter

Official publication of the American Medical Writers Association Pacific Southwest Chapter

POSTSCRIPTS

AIMS AND SCOPE
Postscripts magazine is the official publication of the Pacific Southwest chapter of the American Medical Writers Association (AMWA). It publishes news, notices and authoritative articles of interest in all areas of medical and scientific writing and communications. The scope covers clinical and regulatory writing, scientific writing, publication planning, social media, current regulations, ethical issues, and good writing techniques.

MISSION STATEMENT
The mission of *Postscripts* is to facilitate the professional development of medical writers and serve as a tool to advance networking and mentoring opportunities among all members. Towards this mission, *Postscripts* publishes significant advances in issues, regulations and practice of medical writing and communications; skills and language; summaries and reports of meetings and symposia; and, book and journal summaries. Additionally, to promote career and networking needs of the members, *Postscripts* includes news and event notices covering Chapter activities.

SUBSCRIPTION: *Postscripts* is published monthly except in January and July. The magazine is available as open access publication and is currently distributed online only.

INSTRUCTION FOR CONTRIBUTORS: We consider articles on any topic of interest to medical writers and communicators. It is helpful to look at the past December issues for year end table of contents, and browse past issues for style and type of articles published. We welcome contributions from AMWA members and non-members alike. Please contact editor.

ADVERTISING: *Postscripts* is an advertising-free magazine. However, articles describing products and services relevant to medical writers may be considered or solicited. As a service to our members, they may submit advertisements for their services or products for free. Please contact editor.

WEBSITES:
Postscripts: http://issuu.com/postscripts
Chapter Website: http:// www.amwa-pacsw.org

 Design by Ajay Malik.

Banner Photo Art by Chip Reuben, www.photoartwindows.com / Red phone booth. By Petr Kratochvil, publicdomainpictures.net

POSTSCRIPTS
August 2015 | Volume 5, No.36

Special Focus Issue on Biostatistics. . .

Also in this issue. . .

FEATURES

NETWORKING EVENTS

DEPARTMENTS

COVER:

Jim Yuen, a member of our chapter, shares a picture from his recent trip to Antarctica.

From the President's Desk

"August hangs at the very top of summer, the top of the live-long year, like the highest seat of a Ferris wheel when it pauses in its turning"
Natalie Babbitt, *Tuck Everlasting*

I hope you are enjoying the Summer!

Did you attend our happy hour in La Jolla in June? Or the 3 great medical device presentations in Thousand Oaks in July? Stay tuned for the July meeting summary in our next newsletter. We have more great events planned over the next few months. On August 15th, we are holding a joint meeting with the San Diego Regulatory Affairs Network (SDRAN) about how to write labels that can help improve patient safety. This should be a very informative meeting that also includes a tour of the CareFusion Patient Safety Center.

Are you interested in how medical writers practice their craft? If so, please join us on September 19th at our "Medical Writers' Toolbox Decoded" Symposium. This is a free event hosted by our chapter and Amgen in Thousand Oaks. We thank Ajay Malik, the presenters and the representatives from Amgen for all of their wonderful behind-the-scenes efforts. Be sure to register by September 4th!

In this August newsletter, we send a big thank you to Tanya Hoskin for her very thorough articles about data basics, data types, parametric and nonparametric data and how/why we should look at our data. Dikran Toroser has provided another excellent summary, this time about how to present data in tables and figures.

Are you interested in learning more about grant writing? Please be sure to check out Meg Bouvier's article on how grant writers measure their success. I am sure you will also enjoy the entertaining article by Rebecca Anderson about the different names of commonly-known diseases. We thank April Reynolds for her fun column about what to wear/not to wear to work in the summer, and our employment coordinator, Sharyn Batey, for keeping us informed of jobs in the area.

President-Elect Announced!
I am pleased to announce Susan Vintilla-Friedman, Principal at Vintilla Communications, as the President-Elect of the AMWA Pacific Southwest chapter. Susan has been an active member with our chapter since 2003 and was a member in the Northern California chapter before then. Susan attends our chapter meetings, has presented on our chapter's behalf, is actively involved in the DIA Medical Writing Community, and is on the AMWA Medical Writing Certification Examination Development Committee since September 2013. Susan will become the chapter President in January 2016. Please join me in welcoming Susan to her new role in our chapter leadership!

We also would like to send a big thank you to Andrew Hellman for all of his wonderful support as Secretary of our chapter. Andrew has done a fantastic job in keeping our social media, such as LinkedIn, up to date and also in informing many contacts in industry, academia and sister organizations about our chapter activities. We are sad to see Andrew leave our area but we wish him well in his future. We are happy to announce that in September, Brea Midthune will become our Secretary and Asoka Banno will become our Outreach Coordinator. Thank you to both of them for joining our chapter leadership!

Would you also like to help our chapter? We are looking for an Outreach Coordinator in Thousand Oaks/Los Angeles to help us with planning events in those areas. We always looking for informative and fun articles for our newsletter and it's a great way to get an online writing credit. Please contact Ajay (ajay@amwa-pacsw.org) for more details.

We hope to see you soon!

Donna

Donna Simcoe, MS, MS, MBA, CMPP
President, AMWA Pacific Southwest Chapter

EDITOR'S desk:

Life of Clinical Data: From Creation to Conclusions

Statistics is like taxes. Learning either has a pretty high psychological hurdle. However, once the basic principles are within grasp, they provide a sense of control, confidence, and may help avoid embarrassing pitfalls.

In this issue of *Postscripts*, we reprint a series of 5 articles first published a decade ago by Tanya Hoskin, Senior Statistician at Mayo Clinic. These articles may provide a roadmap from preparing clean reliable datasets to obtaining results and conclusions supported on the strength of the data. We also include 2 articles by Dikran Toroser about presentation of data. Dikran summarizes guidance from the *AMA Manual of Style* on developing Tables and Figures.

The first step in conducting a clinical study or a research project is good planning for data collection, because all downstream activities depend on the completeness of data and logical selection of variables. However, often Medical Writers are invited in the room only after a study is complete, and the first set of data tables becomes available. Nevertheless, this is a very important milestone. At this step, if this data is from a clinical trial, often a Medical Writer may join Clinical Scientist(s), Statistician(s) and folks running the day-to-day operations of a clinical trial, including Clinical Research/Trial Associates to look at the data tables with the goal of cleaning and preparing the dataset for analysis.

Clean Reliable Dataset

In the first article (Data Basics, page 113) of the series, Tanya emphasizes that the time spent on data cleaning is tedious, time-consuming, and unglamorous, but is absolutely essential for obtaining quality data and, thus, robust results and strong conclusions. Using Microsoft Excel spreadsheet as an example, she points to common red flags, including inconsistencies, text where numeric coding should have been used instead or vice versa, or improper handling of missing values. These red flags generate queries to reconcile values with the source or raw data. An audit at this level helps clean and prepare the data for statistical analyses. The goal is to have as clean and as reliable data tables as possible.

The next step in the data cleaning process is to plot and look at the data critically, again—why?—because as the second article in the series (Always Look at the Data, page 115) reminds it is never safe to assume that your dataset is *perfectly* clean. Looking at bars and histograms may reveal obvious errors, improbable range of values, or logical inconsistencies that may have been missed during data cleaning Step 1. Besides helping clean the data further, the patterns help guide the choice of appropriate statistical tests. The third article (Data Types, page 117) is about *knowing* your data. Is the data quantitative or qualitative; if quantitative, is it continuous or discrete; and, if qualitative, is it nominal or ordinal. Why does this *profiling* of data matter? Classifying data into types not only helps flag illogical or inconsistent data, but also help confirm appropriate statistical test for that dataset.

Choosing the Right Statistical Test for the Job

We are not done looking at the data! Now take a step back, survey the landscape, and look at the distribution of data. The next article (More Reasons to Look at the Data, page 121) describes the kind of information that bubbles up when the distribution is assessed, namely, the center, the spread and the shape of the data. Shape helps choose the appropriate statistical procedure. The fifth article in the series (page 124) is a primer about these tests which may be parametric or nonparametric tests. Choosing between parametric and nonparametric tests requires knowledge of subtle differences between these tests, not unlike choosing between *you'll* and *y'all*—while both are contractions of *you all*, likely you will use *you'll* unless you want to give a Southern flavor to your fiction. In brief, know your choice and look smart.

Once, the data has been analyzed, we need to summarize, make presentation and report the statistics properly to support the conclusions drawn from the data. Two articles by Dikran Toroser on page 127 and 129 describe some of the types of tables and figures one can choose from. Next month on Sept 19th at a symposium organized by our Chapter, "Medical Writers' Toolbox Decoded", you can also hear Annalise Nawrocki talk about making Figures and Illustrations (see page 140).

PICTURE: Seattle Daily Times news editor quarters - 1900. Via Wikipedia (http://en.wikipedia.org/wiki/Editing#mediaviewer/File:Seattle_Daily_Times_news_editor_quarters_-_1900.jpg)

Reporting Statistics: The SAMPL Guidelines

Finally, how the statistics are reported in manuscripts, posters and other documents strengthen (or weaken) the validity of conclusions drawn from the data. Several studies have concluded that statistical errors in biomedical literature are mainly due to poor statistical reporting and not due to poor statistical choice of methods.[1] Tom Lang and Douglas Altman, both senior AMWA members, have developed a set of statistical reporting guidelines, "Statistical Analyses and Methods in the Published Literature" or The SAMPL Guidelines. These guidelines are available at the website of EQUATOR network.[2] Further details about statistical reporting are published in a book by these 2 authors.[3]

Increasing the Statistics Literary Quotient

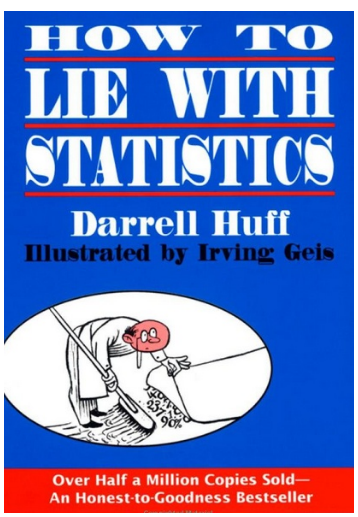
The series of articles on data and statistics presented in this issue of *Postscripts* may serve as a foundation or primer. However, there is much more to learn, including the concepts of inferential statistics, confidence limits, p-values, correlations and regression analysis, etc—these are the towns just coming up on the road taken. Towards the goal of Statistics learning phase 2, there are several textbooks to choose from (visit your local library), free MOOC course,[4] and then there is a classic book from 1954 "How to Lie With Statistics" by Darrell Huff (available at Amazon[5] or as a free ebook fom archive.org.[6]

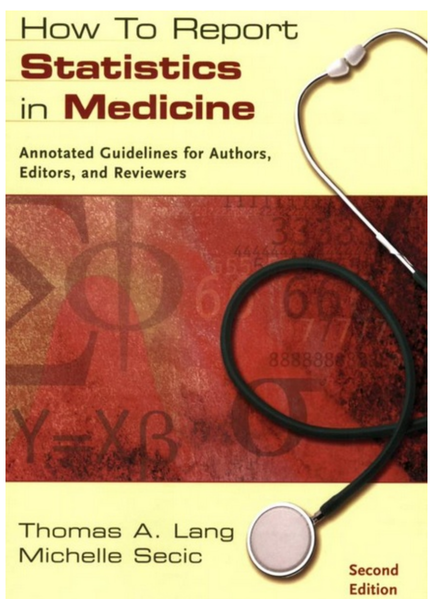
"How to Lie with Statistics"

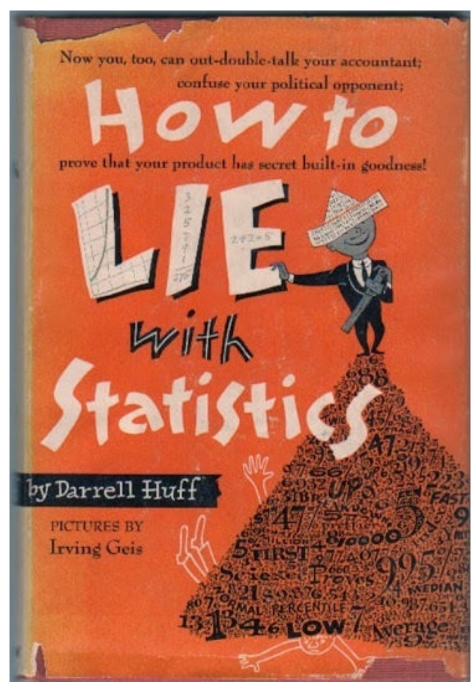
This little book falls in the similar genre as Dr Suess books—entertaining and informative.[7] It is a small book with 142 pages full of examples of fudged analysis or misleading presentation of data, mostly from news media, that will keep the reader smiling and will give no reason to put the book down. Each chapter has several cartoons driving home the statistical concepts and making the reader laugh at the same time! As Darrell admits, the book is devoted to "wilderness of fraud", but in the concluding chapter, he provides tips on "how to recognize sound and usable data in that wilderness of fraud". This book is a *must read* and is a perfect companion to throw in the travel bag while you drive out of your hometown for your summer vacation.

In the end, Medical Writers have everything to gain by being proficient in data analysis and presentation of facts supported by robust statistical analysis. As Darrell Huff wrote:

> "The fact is that, despite its mathematical base, statistics is as much an art as it is science. A great many manipulations and even distortions are possible within the bounds of propriety. Often the statistician must choose among methods, a subjective process, and find the one that he will use to represent the facts."







1. Lang T, Altman D. Basic statistical reporting for articles published in clinical medical journals: the SAMPL Guidelines. In: Smart P, Maisonneuve H, Polderman A (eds). Science Editors' Handbook, European Association of Science Editors, 2013. • Lang TA, Altman DG. Int J Nurs Stud. 2015 Jan;52(1):5-9
2. http://www.equator-network.org/reporting-guidelines/sampl/
3. Lang T, Altman D. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. American College of Physicians. 2nd edition. 2006
4. Statistics in Medicine. Stanford University.https://lagunita.stanford.edu/courses/Medicine/HRP258/Statistics_in_Medicine/about
5. Huff D. How to Lie With Statistics. WW Norton & Co, New York. 1954. http://www.amazon.com/How-Lie-Statistics-Darrell-Huff/dp/0393310728
6. Archive.org. https://archive.org/details/HowToLieWithStatistics
7. Steele JM. Darrell Huff and Fifty Years of How to Lie with Statistics. Statistical Science2005, Vol. 20, No. 3, 205–209. DOI 10.1214/088342305000000205

Data Basics

By Tanya L Hoskin, MS, Senior Statistician

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

Any statistician will tell you that a large percentage of our time is spent cleaning and preparing data for analysis. It is a tedious, time-consuming, and unglamorous part of the job, but it is absolutely essential. The quality of the data determines the quality of our results and the conclusions we draw from them. For the researcher who is collecting, entering, or manipulating data, mastery of fundamental concepts about data is important and can make your and/or your consulting statistician's job easier.

We start with the most fundamental concepts. These ideas are probably best conceptualized by thinking of a spreadsheet, such as Microsoft Excel. Two spreadsheets appear below. Upon first glance, we see that they contain essentially the same information. The structure and quality of the second spreadsheet is far superior to the first, however. From an analysis perspective, the first spreadsheet is unusable in its current form. These two spreadsheets will set the stage for an explanation of the basics of data.

Two fundamental concepts: variables and observational units

- We use the structure of the spreadsheet (columns and rows) to organize our data and enforce consistency. The columns of a spreadsheet represent variables. You could think of a variable as a single piece of information collected on every individual (e.g., height, blood type).
- The rows in the spreadsheet represent the individuals on whom we are collecting data. Typically, it is desirable to have the data for each unique experimental/observational unit (e.g., patient) in a single row of the spreadsheet.

Rules for spreadsheet cells

- Each cell of a spreadsheet (or each variable for an individual) should contain a single piece of information.
- This follows from the fundamental concepts that each column contains a variable (a single piece of information collected on individuals) and each row contains information for a single individual. A cell is the intersection of a particular column and row; therefore, a cell should contain the information for one variable for one individual.
- For example, don't enter gender and date of birth in the same cell. It may seem like a good idea to combine demographic variables or other similar variables, but it's not. Remember that although you can still see the information, software programs cannot easily separate those two pieces of information. You must keep each distinct piece of information in a distinct cell. If you use a comma at any time during your data entry, it's likely a problem!
- See the last column of the "bad spreadsheet" for an illustration of this point. Two different diagnosis codes are entered in a single cell and separated by a comma. See the "good spreadsheet" for one possible solution to this type of situation.

BAD SPREADSHEET

ID	Gender	DOB	Height (cm)	Mass (kg)	Dx
1	M	1/1/1960	163	68	1
2	M	15/1/1961	167	80	2,1
3	F	2/1/25	166	unknown	2
4	MALE	2/15/1963	172cm	82	2
4					3
5	male	March 1, 1964	180	67	2
6	m	3/15/1965	164	64	2 (dx 5/2/00)
7	m	4-1-1966	165 ???	66	1
8	female	April, 1967	166	63	1
9	F	5/1/1968	162	65kg	diabetes
10	f	1969	154	54	2
			average=166		

GOOD SPREADSHEET

ID	Gender (1=M,2=F)	DOB	Height (cm)	Mass (kg)	Dx1	Dx2
1	1	01/01/1960	163	68	1	
2	1	01/15/1961	167	80	2	1
3	2	02/01/1925	166		2	
4	1	02/15/1963	172	82	2	3
5	1	03/01/1964	180	67	2	
6	1	03/15/1965	164	64	2	
7	1	04/01/1966	165	66	1	
8	2	04/15/1967	166	63	1	
9	2	05/01/1968	162	65	3	
10	2	07/01/1969	154	54	2	

Consistency

For data to be useful, it must be recorded in a consistent format. One of the most important tips related to this concept is to avoid using text (letters, symbols, etc.) in your spreadsheet whenever possible. Why? The use of text makes it very difficult to maintain consistency and can

result in a spreadsheet that is very difficult to use for even the simplest analyses.

Whenever possible, you should use numeric coding. For example, if the variable is gender, it is better to develop a numeric coding system such as 1 = male, 2 = female rather than using text such as male/female or M/F. Although this method may seem more cumbersome, think of it this way – there is only one way to write a “1” but many ways to write a text description of gender (“M”, “m”, “male”, etc.). To the software program, “M” and “m” are two different character values. Look at the gender column of the “bad spreadsheet”. If you tried to use this spreadsheet for analysis, the software package would tell you that there are seven different categories for the gender variable. Make certain to carefully document what your numeric codes mean.

- Text added to an otherwise numeric column (for example, including “kg” and entering 50kg for mass) will also make that column very difficult to work with; enter the number only and document the units (in the column name, for example). Refer to the mass column of the “bad spreadsheet” for an example.
- If you need to capture a text description in addition to the variable of interest, feel free to add and use a notes column. Just remember that it is unlikely that you will be able to use this comment field for any type of analysis. Record only numeric values in the column that contains the variable you want to analyze.

Insider information: miscellaneous tips

- There is a difference between a missing value and a zero. Zero is a number and should only be used when a value of zero is observed for the variable of interest. Procedures for entering missing values differ among projects. However, if you are working with a spreadsheet for data entry, leaving the cell blank is often the best choice. Remember that using text such as “N/A” or “unknown” will make an otherwise numeric column difficult to use.
- Use four-digit years for any date variables. Enter dates using a consistent format. The format of date variables should be MM/DD/YYYY. In the “bad spreadsheet”, notice that the use of seven different date formats creates a mess.
- Keep extraneous information, such as data summaries, out of the spreadsheet. The ideal spreadsheet should contain only raw data with a single row of column headings, which contains the variable names. Notice that the last row of the “bad spreadsheet” contains the average height. This is not part of the raw data and should be recorded elsewhere.
- Statistical software packages recognize two variable types: numeric and character. For a variable to be considered numeric, which is desirable in most situations, the entire column must contain only numbers. Statistical software packages read the entire column, and any text in any cell of that column will cause that variable to be treated as character data. For example, the height column in the “bad spreadsheet” would be read in as a character variable. This would make it impossible to get even simple descriptive statistics, such as the mean and standard deviation, without going through special procedures to clean the data.

Although this list cannot cover everything you need to think about when using a spreadsheet to collect data, these concepts and tips should provide a good starting point. If you go back to the “bad spreadsheet” and “good spreadsheet” examples we started with, it should now be clear why the latter is superior.

TANYA L HOSKIN, MS, has worked as a statistician at the Mayo Clinic Deparatment of Health Sciences Research since 2001. She is currently in a senior statistician role collaborating on research projects in disease areas such as fibromyalgia, benign breast disease, and breast cancer and has been a co-author on more than 75 peer-reviewed publications. Previously she worked at her institution’s statistical consulative resource, which included educating clinicians and junior investigators in the basics of statistics and provided the motiviation for writing the series of articles reprinted here. Ms. Hoskin received her Master’s degree in Statistics from Iowa State University.

CREDITS:

Article reprinted with permission of Mayo Foundation for Medical Education and Research.

This article was originally published at mayo.edu in 2009. Visit Biostatistics, Epidemiology and Research Design (BERD) resource center of Mayo Clinic at:
http://www.mayo.edu/ctsa/resources/consultative-resources/biostatistics-epidemiology-and-research-design-berd-resource
—*Editor*

Always Look at the Data

By Tanya L Hoskin, MS, Senior Statistician

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

It is one of the most obvious and yet easily forgotten steps in data analysis – "look at the data." Any consulting statistician will tell you that this is essential to an accurate and appropriate statistical analysis. Why? Well, it is never safe to assume that your data set is perfectly clean. Looking at the data can reveal obvious errors, particularly impossible or improbable values and logical inconsistencies. Furthermore, looking at the data gives you a clearer understanding of the variables and the values they are taking and can help you choose appropriate statistical analyses.

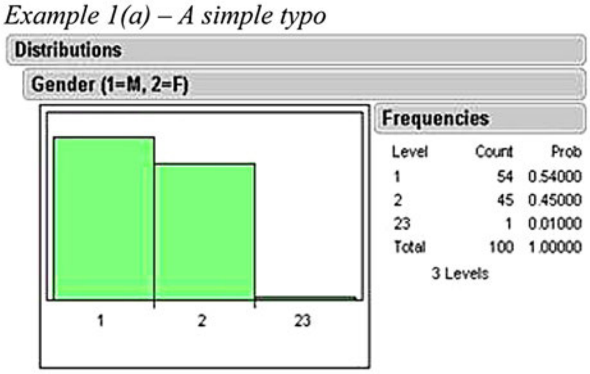
What I mean by "look at the data" is to look at numeric and graphical summaries to identify possible problems in the data and gain descriptive information. Here we will focus on data cleanup. I start by assuming you have a well-organized spreadsheet (see Data Basics; page 113 of this issue). All computer output shown here was generated using JMP 5.0.1 statistical software.

Rule 1: Look at the levels of a categorical variable

Rule 1 deals with variables that classify subjects based on a category. Examples are gender, blood type and histologic stage. For categorical variables, there are usually a small number of possible values. Thus, it makes sense to look at the distribution of these variables (i.e., summarize the categories that occur in the data set) to make certain that each category is valid.

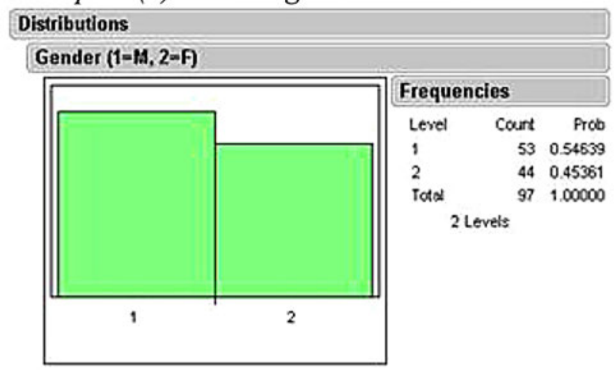
We consider data from a study of 100 patients:

Example 1(a) – A simple typo



The bar chart on the left has one bar for each category or level of the variable. The fact that there are three bars rather than two tells us immediately that there is a problem since gender can only have two possible values. The frequency table on the right also shows us that the levels appearing in the data set are 1, 2 and 23. We need to identify the subject with a 23 recorded, verify the correct gender, and edit the data set accordingly.

Example 1(b) – Missing observations



In this example, the gender variable has the correct number of levels and the correct numeric codes. It would be easy to mistakenly report that 55% of study participants were male and 45% were female. The point of this example is that you should always look at the total sample size included in the statistical procedure. Here the total sample size used was 97 subjects, but there were 100 subjects in the data set. Three patients do not have a value filled in for the gender variable (i.e., missing data). Many times missing data is not an error because it simply reflects reality. For a variable such as gender, however, we should have complete data for all patients. We need to identify the three patients and fill in their gender.

Rule 2 – Look at extreme observations of quantitative variables

Rule 2 deals with variables that can be measured or counted on a numeric scale. Examples of this type of variable are age, systolic blood pressure, total cholesterol, height, weight, and the number of days from hospital admission to discharge. Unlike categorical variables, quantitative variables often have a large number of possible values. Because there are so many possible values, it does not make sense to look at each individual value. We can, however, look at the highest values and the lowest values recorded for a quantitative variable to identify any obvious problems.

We consider the distribution of the age variable in a study of 25 heart attack patients:

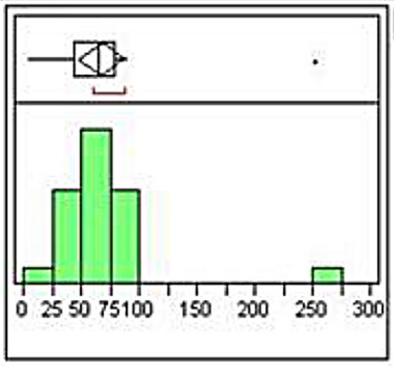
**Example 2 – Multiple problems (*see next page*)

Notice that this display looks different from the displays that we saw in Example 1. The figure on the left is called a histogram. Recall that a bar chart was used to display categorical variables in Example 1, with each bar representing a category. For a quantitative variable such as age or blood pressure, there are no inherent categories. A histogram displays

Example 2 – Multiple problems

Distributions

Age (years)



Quantiles

100.0%	maximum	253.00
99.5%		253.00
97.5%		253.00
90.0%		88.50
75.0%	quartile	79.25
50.0%	median	64.50
25.0%	quartile	44.00
10.0%		35.50
2.5%		4.00
0.5%		4.00
0.0%	minimum	4.00

Moments

Mean	67.916667
Std Dev	44.465831
Std Err Mean	9.0765498
upper 95% Mean	86.69294
lower 95% Mean	49.140393
N	24

quantitative data by grouping the numeric values into intervals. The bar height represents the number of observations falling into that interval. Also notice that the numeric summaries are different. In Example 1, because the variable was categorical, the software summarized the number (frequency) and proportion falling into each category. For a quantitative variable, the software summarizes the mean and standard deviation (under the heading "Moments") as well as the median, minimum, maximum, and other percentiles of the distribution (under the heading "Quantiles").

This histogram shows us that there is a subject with a recorded age above 250 years. Under the heading "Quantiles", we see that the maximum value for age is 253 years, clearly an impossible value. We would need to check this patient's record and correct the data set.

The histogram also shows that there is a patient with a recorded age in the interval from 0 years to 25 years. The quantile display shows us that the minimum age in the data set is 4 years. Although this is not an impossible value, it is an improbable value if this data set contains patients who had a heart attack. It is a good idea to check improbable values to make certain they were recorded correctly.

Finally, we look at the total number of observations included in the calculations. In the display above, in the last row under the heading "Moments", you see that N is 24. Thus, although there were 25 patients in the data set, the age variable was only present for 24. As with gender in Example 1(b), age is a variable for which we should have complete data.

Rule 3 – No one could write down all the rules

Rules 1 and 2 are the basics. Depending on your data set, there could be many additional data checks to perform. Some other data checks to consider:

- Check for duplicate observations. Data cleaning frequently reveals that some patients have multiple records in a data set. In some cases, there is a good reason. In other cases, it is a simple mistake. In either event, you need to be aware so you can handle it appropriately. Consult a statistician if you want to analyze multiple records per patient since this type of analysis requires special methodology.
- Check the order of dates. Date of recurrence should be after date of diagnosis. Date of death cannot be before date of diagnosis. The statements sound obvious, but these types of errors are often missed unless you pay close attention and check. Using software to calculate the time interval between two dates that should have a specific order is an easy way to perform this type of data check. For example, if you calculate the interval between date of death and date of diagnosis, negative values indicate a problem.
- Check logical relationships among variables. If one variable indicates that a patient did not have a CT scan and a second variable has the date of CT scan for that patient as 11/18/2001, the two variables are not consistent. If a data set has gender-specific variables, such as number of pregnancies, these variables should only be filled in for patients of the appropriate gender. The list of logic checks depends on the particular data set. Always take time to think about what logical relationships could be verified in your data.
- Use your medical and subject matter knowledge to identify more subtle inconsistencies.

Collecting your data in a consistent manner with a well-organized spreadsheet is the first step to clean data but not the last. Rare indeed would be a data set that did not suffer from at least a few typos or other mistakes. These rules don't help us find all errors but can help us find the obvious ones. Statisticians regularly check for these types of errors. If you are doing your own simple analyses, make sure you do not neglect this important step.

CREDITS:

Article reprinted with permission of Mayo Foundation for Medical Education and Research.

This article was originally published at mayo.edu in 2009. Visit Biostatistics, Epidemiology and Research Design (BERD) resource center of Mayo Clinic at: http://www.mayo.edu/ctsa/resources/consultative-resources/biostatistics-epidemiology-and-research-design-berd-resource
—Editor

Data Types

By Tanya L Hoskin, MS, Senior Statistician

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

Don't let the title scare you. I know it sounds like one of those topics that only statisticians care about – the kind of topic that makes the eyes of most non-statisticians glaze over. In many ways, data types are very intuitive. However, when you need to collect, record or analyze your data, you can only accomplish these tasks successfully by thinking carefully about what type of data you have.

Suppose an investigator wants to determine if a specific lab value or test result is associated with patient outcome. In most cases, my first question to this investigator would be "What does the variable look like?" In other words, what possible values can the variable take and how will the variable be recorded?

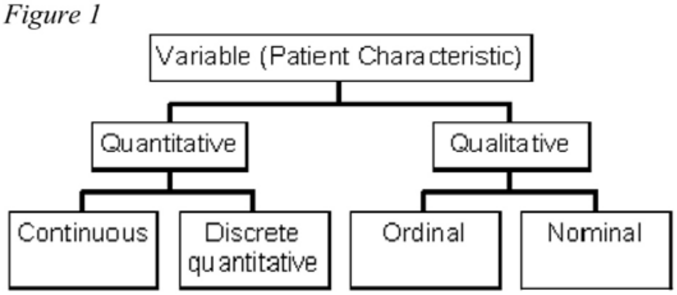
The basics

Typically, a variable can describe either a quantitative or qualitative characteristic of an individual. Examples of quantitative characteristics are age, BMI, creatinine, and time from birth to death. Examples of qualitative characteristics are gender, race, genotype and vital status. Qualitative variables are also called categorical variables. Unfortunately, it gets a little more complicated.

Quantitative and qualitative data types can each be divided into two main categories, as depicted in Figure 1. This means that there are four basic data types that we might need to analyze:

1. Continuous
2. Discrete quantitative
3. Ordinal
4. Nominal

Figure 1



Quantitative variables

You might think of a quantitative variable as one that can only be recorded using a number. These variables describe some quantity about the individual and are often measured (e.g., body mass is measured with a scale) or counted (e.g., the number of needle punctures required to obtain the biopsy specimen is counted).

A quantitative variable can be either continuous or discrete. A continuous variable is one that in theory could take any value in an interval. We say "in theory" simply because we are limited by the precision of the measuring instrument (e.g., a patient's true creatinine value might be 1.21345615 but we might only be able to measure it as 1.213). Examples of continuous variables are body mass, height, blood pressure and cholesterol.

A discrete quantitative variable is one that can only take specific numeric values (rather than any value in an interval), but those numeric values have a clear quantitative interpretation. Examples of discrete quantitative variables are number of needle punctures, number of pregnancies and number of hospitalizations. For these examples, positive whole numbers are the only possible values (i.e., it is not possible to have 1.5 pregnancies).

Qualitative variables

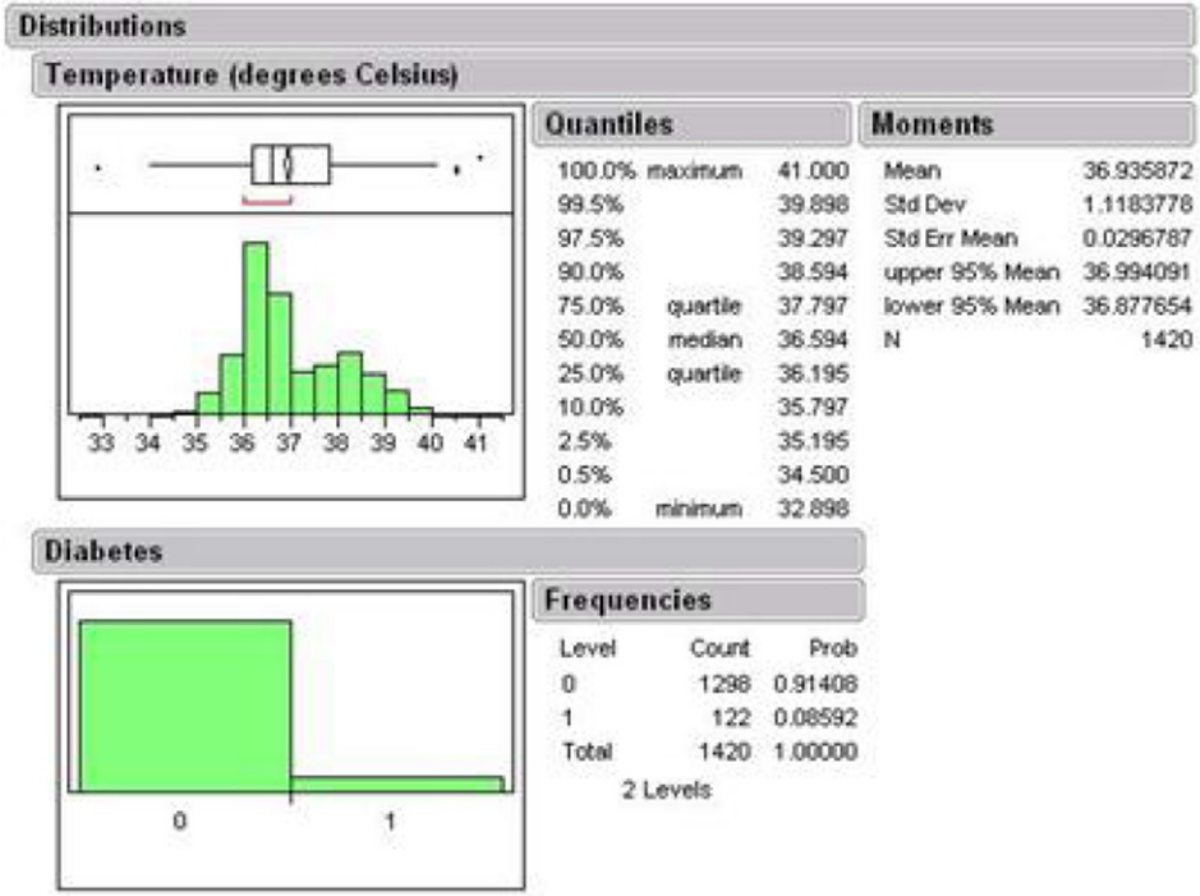
Qualitative or categorical variables describe a quality or attribute of the individual. Categorical data can be either nominal or ordinal. Sex is an example of a nominal variable, and histologic stage is an example of an ordinal variable. What is the difference between these two variables? The values for one of these variables have a specific order; for the other variable, they do not. If one patient has histologic stage 4 and another patient has histologic stage 1, you know that the stage 4 patient has more severe disease. Although the histologic stages are categories, the categories have an inherent order. The same cannot be said for the variable sex. Qualitative data with unordered categories is referred to as nominal; qualitative data with ordered categories is referred to as ordinal.

Why do we care about data types during the data collection phase?

The answer here seems pretty obvious – you must understand the data type of each variable in order to record its values in a consistent manner. This probably won't require much thought in most cases, but consider the following example.

Suppose you are interested in the variable creatinine but plan to analyze it as a binary variable by classifying patients as creatinine < 1.8 or creatinine ³ 1.8. You could simply collect which of these categories each individual falls into, but this probably isn't the best choice. If a categorical variable is based

Figure 2



on the value of a continuous variable, it is generally a good idea to collect the continuous variable. A continuous variable provides more information than a binary variable, which usually translates into more statistical power to detect differences among patients. If, in the analysis phase, you decide that you really do want to use the binary version of the variable, you can easily use a formula in a spreadsheet or statistical software package to create the binary variable from the continuous one you collected. On the other hand, if you only collect the binary variable, you do not have the source measurement recorded to go back to if necessary.

Why do we care about data types during the analysis phase?

You are probably frequently exposed to terms such as mean, median, frequency, proportion, two-sample t-test, chi-square test, regression, correlation, logistic regression, etc. These are all statistical calculations or procedures, but which ones do you use – and when? The appropriate statistical calculation or procedure is driven in large part by the data types.

The most basic example of data types driving statistical calculations is illustrated in Figure 2, which shows the distributions of the variables body temperature (°C) and diabetes (0 = No diabetes, 1 = Yes diabetes) among 1420 hospitalized cancer patients. Diabetes is a nominal variable with only two possible values. Thus, we want to know the number (frequency) of patients with diabetes and what proportion of the total sample they represent. Because body temperature is a continuous variable with many possible values, we summarize its distribution by reporting statistics such as the median, minimum, maximum, mean and standard deviation. Clearly it would not be feasible or helpful to summarize the number and proportion of patients who had each specific body temperature value, just as it would make no sense to calculate the mean of the diabetes variable.

An analysis example

When you begin to pursue analyses more complex than descriptive statistics, data types are just as important and will lead you to the appropriate statistical procedures. Consider two questions in a hypothetical study involving these 1420 patients:

1. Is gender associated with body temperature?
2. Is gender associated with diabetes?

First, we must consider the data types. We know that body temperature is quantitative and more specifically

continuous; diabetes is qualitative and more specifically nominal. Okay, now what? What information might be helpful in addressing the first question? Would it be helpful to know the distribution of body temperature separately for males and females?

The distributions shown in Figure 3 summarize a continuous variable (body temperature) for each of two groups (females and males). A statistical quantity used to summarize the distribution of a continuous variable is the mean. We see that the mean body temperature for males was 36.90°, compared to 36.99° for females. Just as we compare means in the two groups in our descriptive statistical analysis, we need a procedure that will statistically compare the mean among males to the mean among females.

One statistical test for comparing means between two groups is a two-sample t-test.

To answer question 2, we might start by summarizing the distribution of the diabetes variable separately for males and females (Figure 4).

The statistical quantity used to summarize the distribution of a nominal variable such as diabetes is a proportion. From Figure 4 we see that 46/562 (8.2 percent) females have diabetes, compared to 76/858 (8.9 percent) males. Because of the data types, we know that we would need a statistical procedure to compare proportions. The appropriate procedure to statistically compare proportions between two groups is a chi-square or Fisher's exact test.

Figure 3

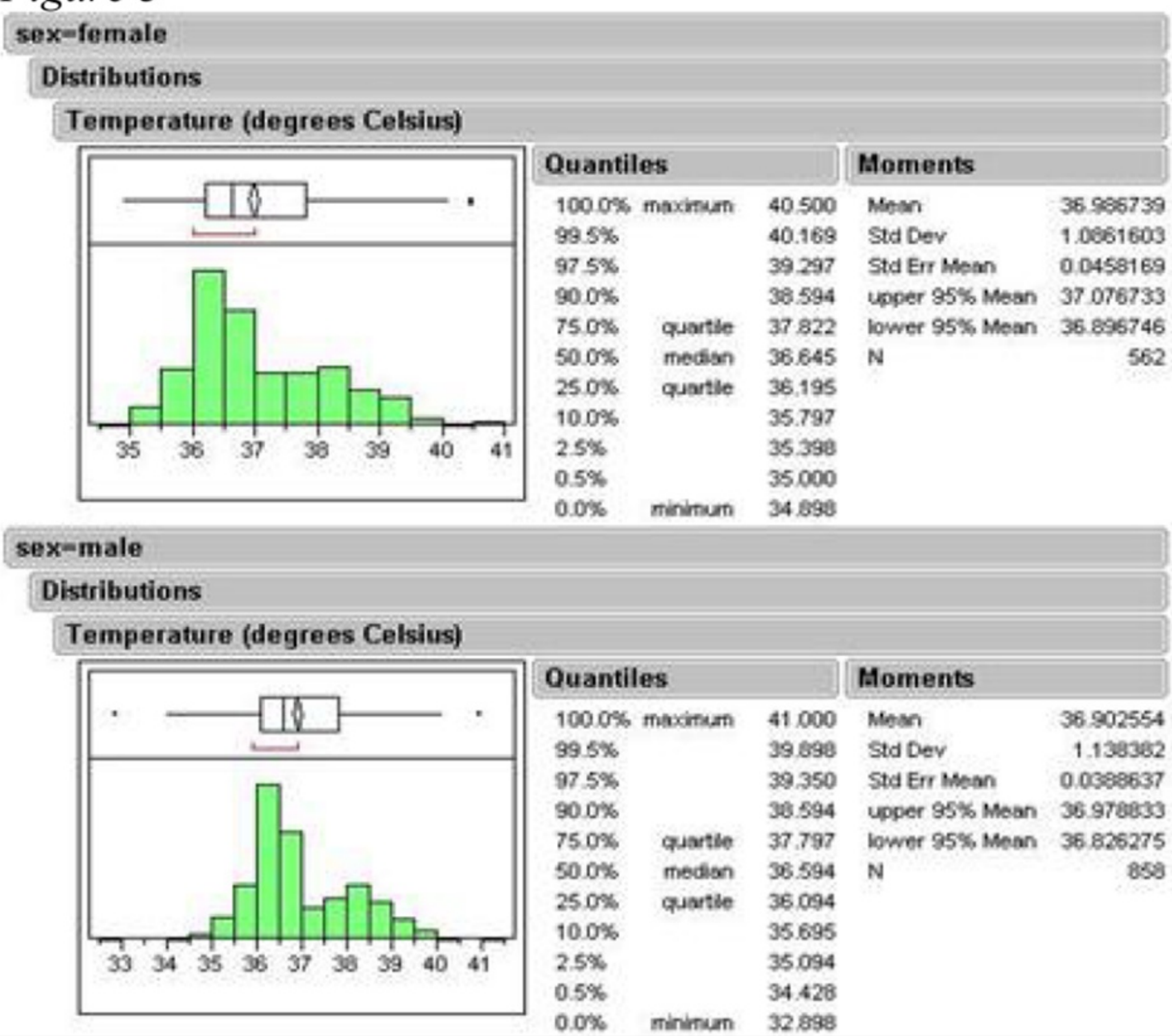


Figure 4

Conclusion

It's not always easy to classify the data type of a variable or to decide how it should be analyzed. Continuous and nominal variables are usually straightforward, but discrete quantitative and ordinal variables can be more challenging. For example, if you are interested in reporting the number of pregnancies among women in your study group, is it meaningful to treat this as a continuous variable and provide the mean number of pregnancies? Or would it be more meaningful to treat it as an ordinal variable and summarize the number of women with one pregnancy, two pregnancies, etc.? Or would it be more meaningful to report the number of women who had one or more pregnancies?

The answers to questions like these often depend on many factors such as the reason that you are summarizing a particular variable and what you believe will be the most meaningful and useful statistic for your audience.

CREDITS:

This article was originally published at mayo.edu in 2009. Visit Biostatistics, Epidemiology and Research Design (BERD) resource center of Mayo Clinic at: http://www.mayo.edu/ctsa/resources/consultative-resources/biostatistics-epidemiology-and-research-design-berd-resource
—*Editor*

More Good Reasons to Look at the Data

By Tanya L Hoskin, MS, Senior Statistician

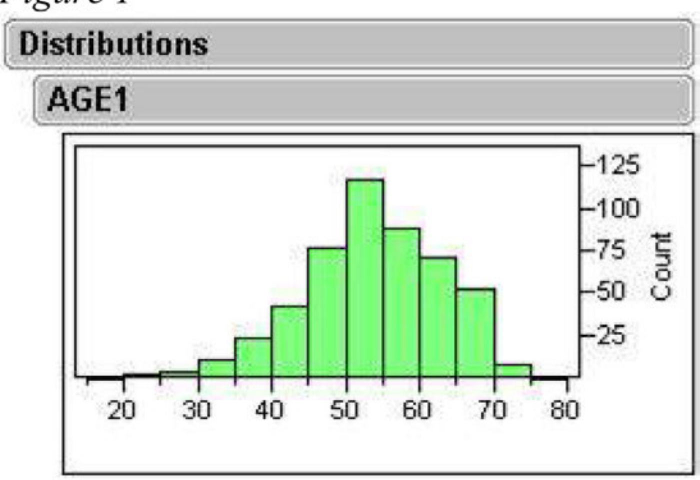
Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

The article *Always Look at the Data* (see page 115) suggested that you look at your data to find problems, such as invalid codes or impossible values. Another article, *Data Types* (see page 117), explained how the type of data helps determine what statistical procedures are appropriate. This statistical tip will be a combination of those two topics. Here, we will think about how the overall look of our data might also help us determine which statistical procedures to use. In other words, we are going to think about aspects of the data distribution. We will focus on quantitative, continuous data.

What is a distribution?

You might hear the word "distribution" all the time in discussions about data, but how often do you think about what it means? Basically, we have a clinical question we want to answer. In order to answer that question, we collect data from a sample of individuals from the population. It is not very helpful to look at each individual's value separately, so we need a way to look at the values for the whole sample at once. The distribution of a variable shows us a summary of all the values in a single picture. In other words, the distribution shows us how the values are distributed or where they fall in the range of possible values.

Figure 1



The histogram in Figure 1 shows the distribution of the age variable (AGE1) for a sample of 500 individuals. The vertical bars represent the number of patients out of the total sample who have ages in each interval. Here the intervals span 5 years. The tallest bar shows us that approximately 120 of the individuals had an age between 50 and 55 years. This bar is approximately in the middle of the range of values. Without calculating the average age of our patients, we might guess that it is between 50 and 55 just by looking at this picture. We also know that most patients were between 30 and 70 years old. There were only a few patients younger than 30 or older than 70 based on the very short bars for those intervals. Clearly, we gain a lot of information quickly by looking at the distribution, and this information is certainly more helpful than a list containing each individual's age.

Center, spread and shape

The center, spread and shape of a data distribution are the three key pieces of information we can assess by looking at a histogram. "Center" simply refers to the middle of the distribution, or an estimate of what a typical value would be for these individuals. Based on Figure 1, we said that the center of the distribution seems to be somewhere between 50 and 55. "Spread" is simply how "spread out" or variable the data is. In other words, were a wide range of values observed or do patients generally have values near e center of the distribution?

Figure 2

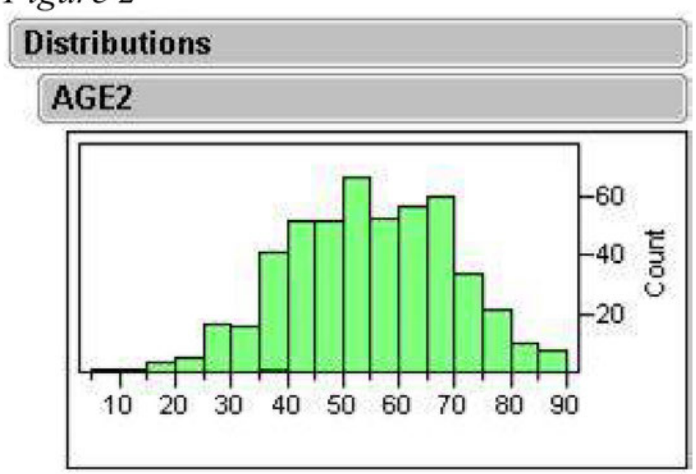


Figure 2 contains the distribution of ages (AGE2) for a different sample of 500 patients. The distribution in Figure 2 has more spread than the distribution in Figure 1. Can you see why? Well, the center of the distribution is about the same, somewhere between 50 and 55, but we observed a larger range of values. The youngest patient in Figure 2 is between 5 and 10 years of age, while the oldest is between 85 and 90. Also, we see more patients in intervals that are further from the center (e.g., 25 to 30 years). These observations tell us that there is more variability or spread in age among the patients from Figure 2 compared to Figure 1.

If you were to describe the shape of the histograms in Figures 1 and 2, you might say that both are shaped like a mound. The general appearance of the histogram is one thing to note about the shape of the distribution; in many cases it will be a mound, but occasionally you might see a shape that looks like two mounds (called a bimodal distribution) or some other non-mound shape. A second very important property to assess with regard to the shape of a distribution is its symmetry. In other words, could you cut the histogram in half and have one side that looked roughly like the other side? If a distribution is not symmetric, we might describe it as either right or left skewed.

Figure 3 shows two examples of distributions with skewed shapes. The distribution of the variable ejection fraction (EF) could be described as left skewed; the distribution of the variable heart rate could be described as right skewed. The “right” and “left” parts of the description refer to the side of the histogram that trails out farther than the other side.

Shape helps determine appropriate statistical procedures: a simple example

Look at the distribution of creatinine in Figure 4. It is based on a sample of 20 individuals and clearly has a right skew. The sample mean and median are 1.5 and 1.2, respectively. In this example, these two measures of center are quite different. Which one should you report? Which value is more representative of a “typical value” for these individuals?

Figure 3

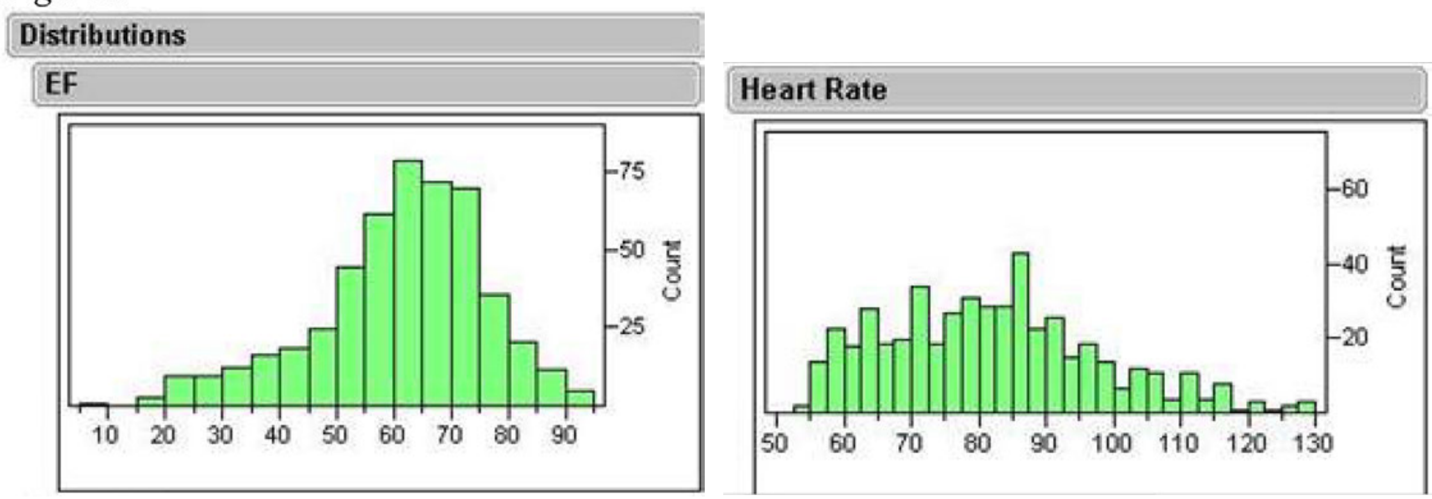
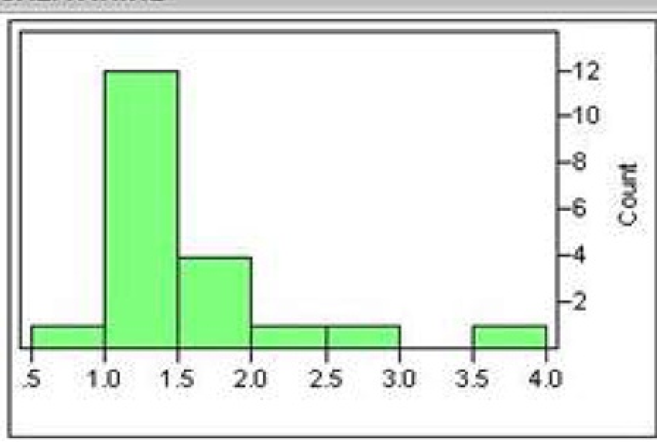


Figure 4

Distributions

CREATININE



Quantiles

100.0%	maximum	3.7998
99.5%		3.7998
97.5%		3.7998
90.0%		2.8300
75.0%	quartile	1.7000
50.0%	median	1.2499
25.0%	quartile	1.0999
10.0%		1.0000
2.5%		0.8999
0.5%		0.8999
0.0%	minimum	0.8999

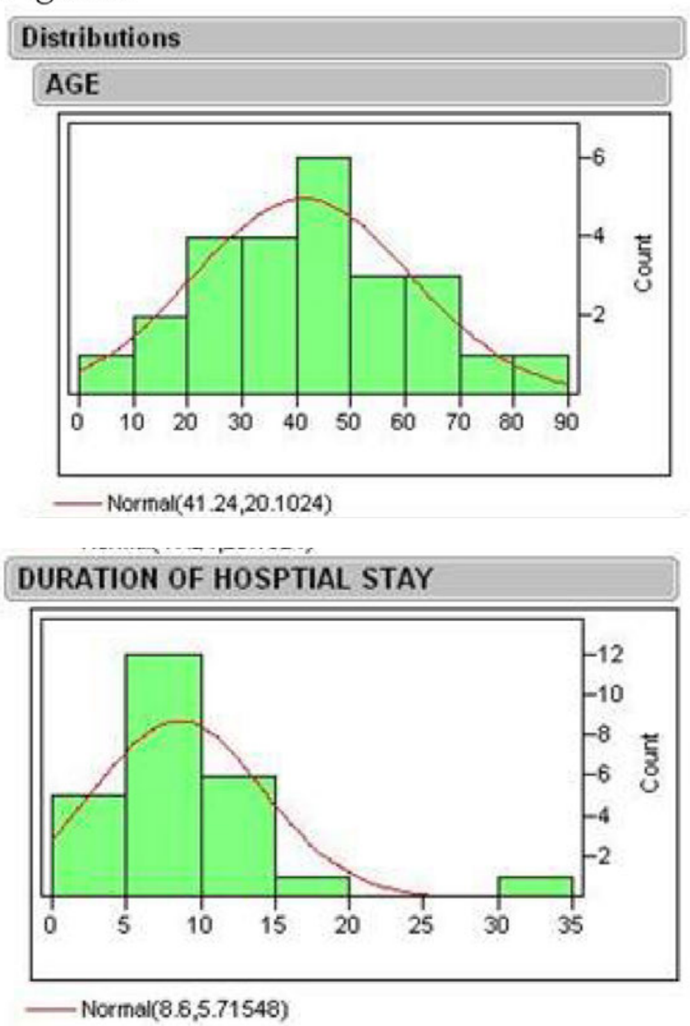
Moments

Mean	1.544917
Std Dev	0.7192379
Std Err Mean	0.1608265
upper 95% Mean	1.8815307
lower 95% Mean	1.2083033
N	20

For highly skewed distributions or those with unusually large or small values (i.e., outliers), the median typically is a more appropriate summary statistic to describe the center of the distribution. Why? Since the mean is calculated by adding all of the individual values and dividing by the sample size, it is strongly influenced by extreme observations in the tails of the distribution, especially for small sample sizes. The median, on the other hand, is simply the value that has half of the data points falling below it and half above, so it is not affected by the magnitude of extreme observations. In a perfectly symmetric distribution, the mean and median are equal. In a skewed distribution, the mean is pulled in the direction of the skew. Thus, the mean is larger than the median if the distribution is right skewed and smaller than the median if the distribution is left skewed.

When one reports the median rather than the mean, the range (minimum and maximum) or interquartile range (25th and 75th percentiles) is the appropriate measure of spread. For the distribution of creatinine in Figure 4, we might report a median creatinine of 1.2 with a range from 0.9 to 3.8.

Figure 5



The normal distribution

In order to talk about more advanced examples of how the shape of the data distribution affects our choice of statistical procedures, we must discuss the normal probability distribution. There are assumptions underlying many common statistical tests and procedures, and the most common assumptions are those related to the normal probability distribution. Specifically, many statistical procedures assume your data is normally distributed.

The normal probability distribution is a theoretical, mathematically defined distribution that explicitly defines how likely certain ranges of values are to occur. It is perfectly symmetric and shaped like a bell. A normal probability distribution curve overlays the two histograms in Figure 5. You can see that the histogram more closely resembles the shape of the normal curve for the variable age than for the variable duration of hospital stay. We might say that the age distribution is approximately normal but that duration of hospital stay is not normal and is right skewed.

The normal distribution shows up in a surprising number of natural phenomena. For example, human hippocampal volumes are approximately normal. It also shows up in many of our derived statistical calculations. By exploiting the desirable properties of the normal distribution, statisticians have developed many of the statistical procedures that are so helpful for answering scientific questions.

If the normal distribution is at the foundation of so many of our statistical procedures, you might wonder how we deal with the fact that the histograms are rarely, if ever, perfectly normal. We will discuss this topic in the next quarterly statistical tip, but the short answer is that “approximately normal” is often good enough, and we have tools to help when distributions are clearly not normal.

CREDITS:

This article was originally published at mayo.edu in 2009. Visit Biostatistics, Epidemiology and Research Design (BERD) resource center of Mayo Clinic at: http://www.mayo.edu/ctsa/resources/consultative-resources/biostatistics-epidemiology-and-research-design-berd-resource
—*Editor*

Parametric and Nonparametric: Demystifying the Terms

By Tanya L Hoskin, MS, Senior Statistician

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

In *More Good Reasons to Look at the Data* (see page 121), we looked at data distributions to assess center, shape and spread and described how the validity of many statistical procedures relies on an assumption of approximate normality. But what do we do if our data are not normal? In this article, we'll cover the difference between parametric and nonparametric procedures. Nonparametric procedures are one possible solution to handle non-normal data.

Definitions

If you've ever discussed an analysis plan with a statistician, you've probably heard the term "nonparametric" but may not have understood what it means. Parametric and nonparametric are two broad classifications of statistical procedures. *The Handbook of Nonparametric Statistics*[1] from 1962 (p. 2) says:

> "A precise and universally acceptable definition of the term 'nonparametric' is not presently available. The viewpoint adopted in this handbook is that a statistical procedure is of a nonparametric type if it has properties which are satisfied to a reasonable approximation when some assumptions that are at least of a moderately general nature hold."

That definition is not helpful in the least, but it underscores the fact that it is difficult to specifically define the term "nonparametric." It is generally easier to list examples of each type of procedure (parametric and nonparametric) than to define the terms themselves. For most practical purposes, however, one might define nonparametric statistical procedures as a class of statistical procedures that do not rely on assumptions about the shape or form of the probability distribution from which the data were drawn.

The short explanation

Several fundamental statistical concepts are helpful prerequisite knowledge for fully understanding the terms "parametric" and "nonparametric." These statistical fundamentals include random variables, probability distributions, parameters, population, sample, sampling distributions and the Central Limit Theorem. I cannot explain these topics in a few paragraphs, as they would usually comprise two or three chapters in a statistics textbook. Thus, I will limit my explanation to a few helpful (I hope) links among terms.

The field of statistics exists because it is usually impossible to collect data from all individuals of interest (population). Our only solution is to collect data from a subset (sample) of the individuals of interest, but our real desire is to know the "truth" about the population. Quantities such as means, standard deviations and proportions are all important values and are called "parameters" when we are talking about a population. Since we usually cannot get data from the whole population, we cannot know the values of the parameters for that population. We can, however, calculate estimates of these quantities for our sample. When they are calculated from sample data, these quantities are called "statistics." A statistic estimates a parameter.

Parametric statistical procedures rely on assumptions about the shape of the distribution (i.e., assume a normal distribution) in the underlying population and about the form or parameters (i.e., means and standard deviations) of the assumed distribution. Nonparametric statistical procedures rely on no or few assumptions about the shape or parameters of the population distribution from which the sample was drawn.

As I mentioned, it is sometimes easier to list examples of each type of procedure than to define the terms. Table 1 contains the names of several statistical procedures you might be familiar with and categorizes each one as parametric or nonparametric. All of the parametric procedures listed in Table 1 rely on an assumption of approximate normality.

An example

Suppose you have a sample of critically ill patients. The sample contains 20 female patients and 19 male patients. The variable of interest is hospital length of stay (LOS) in days, and you would like to compare females and males. The histograms of the LOS variable for males and females appear in Figure 1. We see that the distribution for females has a strong right skew. Notice that the mean for females is 60 days while the median is 31.5 days. For males, the distribution is more symmetric with a mean and median of 30.9 days and 30 days, respectively. Comparing the two groups, their medians are quite similar, but their means are very different. This is a case where the assumption of normality associated with a parametric test is probably not reasonable. A nonparametric procedure would be more appropriate.

Table 1

Analysis Type	Example	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Is the mean systolic blood pressure (at baseline) for patients assigned to placebo different from the mean for patients assigned to the treatment group?	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Was there a significant change in systolic blood pressure between baseline and the six-month follow-up measurement in the treatment group?	Paired t-test	Wilcoxon signed-rank test
Compare means between three or more distinct/independent groups	If our experiment had three groups (e.g., placebo, new drug #1, new drug #2), we might want to know whether the mean systolic blood pressure at baseline differed among the three groups?	Analysis of variance (ANOVA)	Kruskal-Wallis test
Estimate the degree of association between two quantitative variables	Is systolic blood pressure associated with the patient's age?	Pearson coefficient of correlation	Spearman's rank correlation

Table 1. A listing of parametric tests and analogous nonparametric procedures

This is the situation listed in the first row of Table 1 – comparing means between two distinct groups. Thus, the appropriate nonparametric procedure is a Wilcoxon rank-sum test. This test would give us a p-value of 0.63. Those of you familiar with p-values know that we typically compare our p-value to the value 0.05. We usually say that a p-value less than 0.05 is an indication of a statistically significant result. So, we would say that there is no significant difference between the genders with respect to length of stay based on the Wilcoxon rank-sum test. Incidentally, the p-value for the two-sample t-test, which is the parametric procedure that assumes approximate normality, is 0.04. You can see that in certain situations parametric procedures can give a misleading result.

Why don't we always use nonparametric tests?

Although nonparametric tests have the very desirable property of making fewer assumptions about the distribution of measurements in the population from which we drew our sample, they have two main drawbacks. The first is that they generally are less statistically powerful than the analogous parametric procedure when the data truly are approximately normal. "Less powerful" means that there is a smaller probability that the procedure will tell us that two variables are associated with each other when they in fact truly are associated. If you are planning a study and trying to determine how many patients to include, a nonparametric test will require a slightly larger sample size to have the same power as the corresponding parametric test.

The second drawback associated with nonparametric tests is that their results are often less easy to interpret than the results of parametric tests. Many nonparametric tests use rankings of the values in the data rather than using the actual data. Knowing that the difference in mean ranks between two groups is five does not really help our intuitive understanding of the data. On the other hand, knowing that the mean systolic blood pressure of patients taking the new drug was five mmHg lower than the mean systolic blood pressure of patients on the standard treatment is both intuitive and useful.

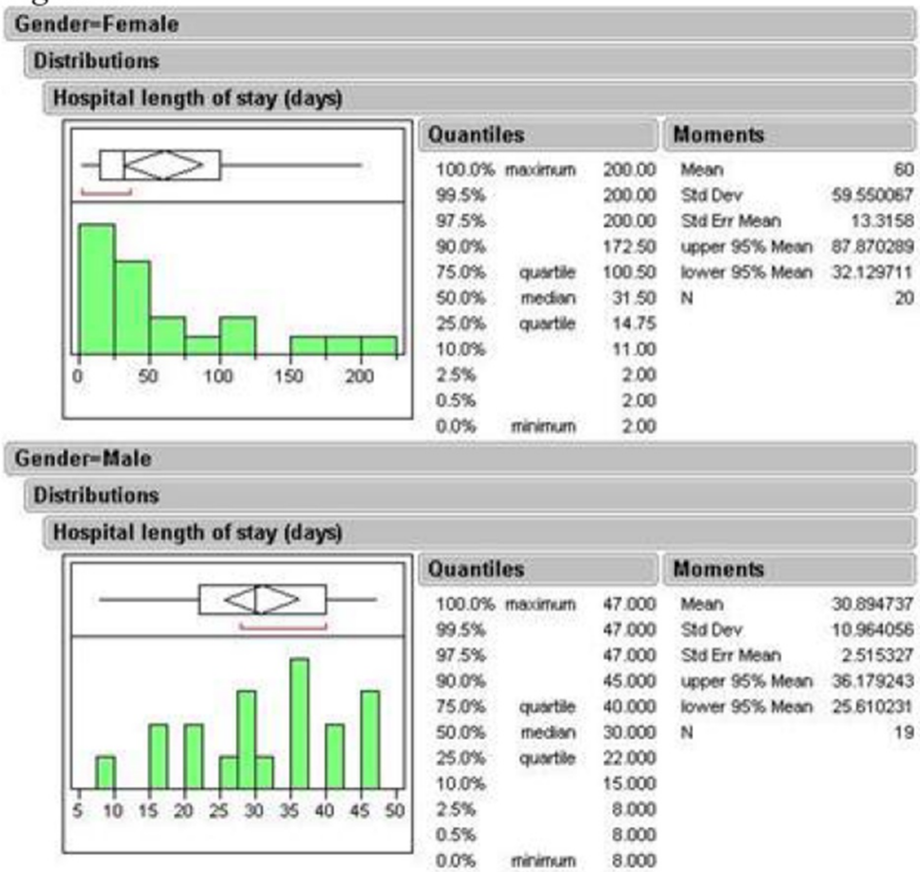
In short, nonparametric procedures are useful in many cases and necessary in some, but they are not a perfect solution.

Take-home points

Here is a summary of the major points and how they might affect statistical analyses you perform:

- Parametric and nonparametric are two broad classifications of statistical procedures.
- Parametric tests are based on assumptions about the distribution of the underlying population from which the sample was taken. The most common parametric assumption is that data are approximately normally distributed.

Figure 1



- Nonparametric tests do not rely on assumptions about the shape or parameters of the underlying population distribution.
- If the data deviate strongly from the assumptions of a parametric procedure, using the parametric procedure could lead to incorrect conclusions.
- You should be aware of the assumptions associated with a parametric procedure and should learn methods to evaluate the validity of those assumptions.
- If you determine that the assumptions of the parametric procedure are not valid, use an analogous nonparametric procedure instead.
- The parametric assumption of normality is particularly worrisome for small sample sizes ($n < 30$). Nonparametric tests are often a good option for these data.
- It can be difficult to decide whether to use a parametric or nonparametric procedure in some cases. Nonparametric procedures generally have less power for the same sample size than the corresponding parametric procedure if the data truly are normal. Interpretation of nonparametric procedures can also be more difficult than for parametric procedures.
- Visit with a statistician if you are in doubt about whether parametric or nonparametric procedures are more appropriate for your data.
- The book Practical Nonparametric Statistics[2] is an excellent resource for anyone interested in learning about this topic in great detail. More general texts such as Fundamentals of Biostatistics[3] and Intuitive Biostatistics[4] have chapters covering the topic of nonparametric procedures.

References

1. Walsh, J.E. (1962) Handbook of Nonparametric Statistics, New York: D.V. Nostrand.
2. Conover, W.J. (1980). Practical Nonparametric Statistics, New York: Wiley & Sons.
3. Rosner, B. (2000). Fundamentals of Biostatistics, California: Duxbury Press.
4. Motulsky, H. (1995). Intuitive Biostatistics, New York: Oxford University Press.

CREDITS:

This article was originally published at mayo.edu in 2009. Visit Biostatistics, Epidemiology and Research Design (BERD) resource center of Mayo Clinic at: http://www.mayo.edu/ctsa/resources/consultative-resources/biostatistics-epidemiology-and-research-design-berd-resource
—*Editor*

AMA-zing Style — the AMA Manual of Style Column

By Dikran Toroser, PhD, CMPP, Amgen Inc., Thousand Oaks, Calif.

Presentation of Data: Tables

When planning your publication, you will have to consider the best way to communicate information to your audience. Generally, data summaries may take the form of text, tables or figures. Tables are a convenient way to present lists of numbers or text in columns and are ideal to explain variables. Tables can also make an article more readable by removing numeric or listed data from the text. *The AMA Manual of Style* contains excellent information on table design.

It is just as important to think about the organization of tables as it is to think about the organization of paragraphs. A well-organized table allows readers to grasp the meaning of the data presented with ease, while a disorganized one will leave the reader confused about the data itself, or the significance of the data.

In terms of space, a well-structured table is one of the most efficient ways to convey a large amount of data. As text, the information may take much more space; and as a figure, details and precise values may be less apparent. Priorities in the creation and publication of tables are to emphasize important information efficiently and to ensure that each table makes a clear point. In addition to presenting study results, tables can be used to explain or amplify the methods or highlight other key points in the article. Like a paragraph, each table should be cohesive and focused.

Although, tabular presentation of data adds variety to the article layout, authors and editors should avoid using tables and figures simply to break up text or to impart visual interest. Below are some considerations when making the choice between text, table or figure to present your data.

Types of Tables:

Regular table. Displays information arranged in columns and rows and is used most commonly to present numerical data. Tabulating *all* collected study data is unnecessary and actually distracts and overwhelms the reader. Tables should be able to stand independently, without requiring explanation from the text.

Tabulation. This is a brief, in-text table that may be used to set material off from text. Tabulations require the text to explain their meaning.

Guidelines for Using Text *vs* Tables *vs* Figures to Display Data

Uses of Text

- Present quantitative data that can be given concisely and clearly
- Describe *simple* relationships among data

Uses of Tables

- Present large amounts of quantitative information in a smaller space
- Demonstrate detailed item-to-item comparisons
- Display many quantitative values simultaneously
- Display individual data values precisely
- Demonstrate complex relationships in data

Uses of Figures

- Highlight patterns or trends in data
- Demonstrate changes or differences over time
- Display complex relationships among quantitative variables
- Clarify or explain methods
- Provide information to enhance understanding of complex concepts
- Provide visual data to illustrate findings (eg, slides, photographs, maps)
- Illustrate scientific or clinical concepts, mechanisms, or pathophysiology

Matrix. This is a tabular structure that uses numbers, short words (eg, no, yes), or symbols (eg, bullets, check marks) to depict relationships among items in columns and rows.

Organizing Information in Tables. Tables should be organized into columns and rows by type and category, thereby simplifying access and display of data and information. During planning and creation, the writer should consider the primary comparisons of interest. The primary comparisons should be shown horizontally across the table.

Table Components:
Tables usually contain 5 major elements: title, column headings, stubs (row headings), body (data field) consisting of individual cells (data points), and footnotes.

Titles should be succinct, specific and written as a phrase rather than as a sentence.

Column Headings. Main categories in the table should have separate columns—each with a brief heading. In tables for studies that have independent and dependent variables, the independent variables conventionally are displayed in the left-hand column (stub) and the dependent variables in the columns to the right.

Table Stubs (Row Headings). The left-most column of a table contains the table stubs (or row headings), which are used to label the rows of the table and apply to all items in that row.

Field. The field or body of the table presents the data. Each data entry point is contained in a cell, which is the intersection of a column and a row.

Totals. Any discrepancies in the totals (eg, because of rounding) should be explained in a footnote. Boldface should **not** be used to overemphasize data in the table (eg, significant odds ratios or *P*-values).

Rules and Shading. For *JAMA* and the *Archives* Journals, tables should be submitted without rules drawn in (as opposed to table borders, which are appropriate) or shading. Many journals add rules and shading during the production process. For example, *JAMA* uses horizontal rules to separate rows of data. Other journals use shading for the same purpose.

Footnotes. The order of the footnotes is determined by the placement in the table of the item to which the footnote refers. The letter for a footnote that applies to the entire table (eg, one that explains the method used to gather the data or format of data presentation) should be placed after the table title. A footnote that applies to 1 or 2 columns or rows should be placed after the column heading(s) or stub(s) to which it refers. A footnote that applies to a single entry in the table or to several individual entries should be placed at the end of each entry to which it applies. For both tables and figures, footnotes are indicated with superscript lowercase letters in alphabetical order (a-z). The font size of the footnote letters should be large enough to see clearly without appearing to be part of the actual data.

Units of Measure. In tables, units of measure, including the variability of the measurement if reported, should follow a comma in the table column heading or stub. The following are examples of stub entries with units of measure:

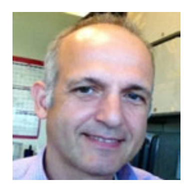
Age, mean (SD), y
Body mass index, median (IQR)

Abbreviations. Within the table, units of measure may be abbreviated for space considerations. However, spelled-out words should ***not*** be combined with abbreviations for units of measure. "1st wk" or "Week 1" is acceptable, but not "First wk."

Numbers. Additional digits (including zeros) should not be added, eg, after the decimal point, to provide all data entries with the same number of digits. Doing so may indicate more precise results than actually were calculated or measured. Values for reporting statistical data, such as *P* values and confidence intervals, also should be presented and rounded appropriately.

Guidelines for Preparing and Submitting Tables. Authors submitting tables in a scientific article should consult the publication's instructions for authors for specific requirements and preferences regarding table format.

See pages 81-98 in the AMA Manual of Style 10th edition for additional information.



DIKRAN TOROSER, PhD, CMPP, a member of the AMWA Pacific Southwest chapter, is a regular contributor to the *Postscripts* magazine since 2012. He developed the monthly AMAzing Style column which covers topics from the AMA Manual of Style, and has also written on publication-related topics in these pages. Dikran is currently a Senior Medical Writing Manager at Amgen Inc. in Thousand Oaks, California. He earned his PhD in Biochemistry from Newcastle University (UK), and did his post-doctoral training in biochemical genetics at the John Innes Center of the Cambridge Laboratory (Norwich, UK) and in molecular biology with the USDA. Prior to Amgen, Dikran was on the faculty (research) at the School of Pharmacy at the University of Southern California. He can be reached at dtoroser@amgen.com.

AMA-zing Style — the AMA Manual of Style Column

By Dikran Toroser, PhD, CMPP, Amgen Inc., Thousand Oaks, Calif.

Types of FIGURES

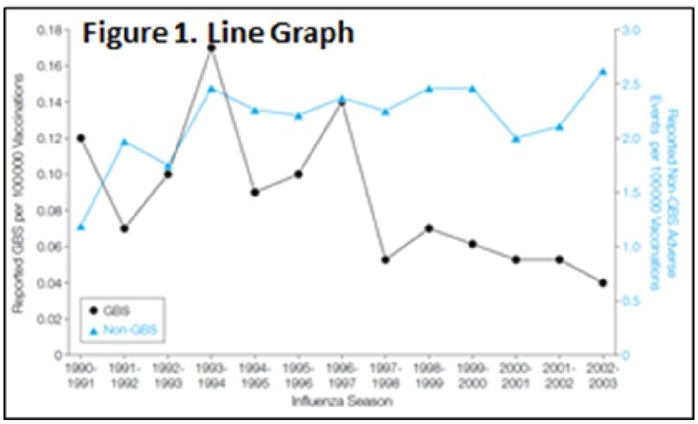
Communication of data requires figures in addition to words and equations. If a graph is appropriate, you need to make some deliberate choices. The AMA manual contains invaluable information on your available choices.

The term *figure* refers to any graphical display used to present information or data, including statistical graphs, maps, algorithms, illustrations, computer generated images, and photographs. In scientific articles, selection of a particular type of figure depends on the purpose and type of information being displayed. Some of the most common types of figures in biomedical publications are discussed below.

Statistical Graphs

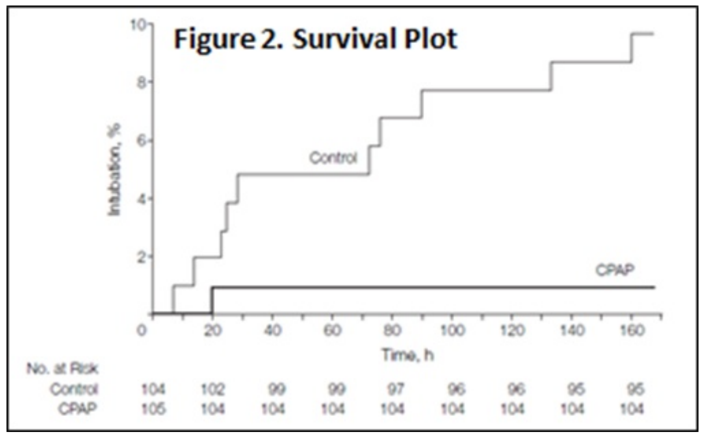
Line Graphs. These have 2 or 3 axes with continuous quantitative scales demonstrating the relationship between 2 or more variables, such as changes over time. Usually, the dependent variable is on the vertical axis (y-axis) and the independent variable on the horizontal axis (x axis) (Figure 1).

Figure 1 Line Graph (adapted from Haber et al (2004) JAMA, 292(20): 2478-2481)



Survival Plots. These plots, of time-to-event outcomes, such as from Kaplan-Meier analyses, display the proportion of individuals, represented on the y-axis as a proportion or percentage, remaining free of or experiencing a specific outcome over time, represented on the x-axis. When the outcome of interest is relatively frequent, event-free survival is plotted on the y-axis from 0 to 1.0 (or 0% to 100%), with the curve starting at 1.0 (100%). When the outcome is relatively infrequent (occurs in < 30% of the study population), it is preferable to plot upward starting at 0 so that the curves can be seen without breaking or truncating the y-axis scale. The curve should be drawn as a step function (not smoothed). The number of individuals followed up for each time interval (number at risk) should be shown underneath the x-axis. Time-to-event estimates become less certain as the number of individuals diminishes, so consideration should be given to not displaying data when less than 20% of the study population is still in follow-up.

Figure 2 Survival Plot (adapted from Squadrone et al JAMA, 2005; 293(5): 589)



Scatterplots. In scatterplots, individual data points are plotted according to coordinate values with continuous, quantitative x- and y-axis scales. By convention, independent variables are plotted on the x-axis and dependent variables on the y-axis. Data markers are not connected by a curve, but a curve that is generated mathematically may be fitted to the data and summarize the relationship among the variables. The statistical method used to generate the curve and the statistic that summarizes the relationship between the dependent and independent variables, such as a correlation or regression coefficient, should be provided in the figure or legend (Figure 3).

Bar Graphs. Bar graphs have a single axis and are used to display frequencies (counts or percentages) on the axis according to categories shown on a baseline. A bar graph is typically vertical, with frequencies shown on a vertical y-axis, but may be horizontal. Data in each category are represented by a bar. Bars should have the same width, be separated by a space, and be wider than the space

between them. Bar lengths are proportional to frequency, the scale on the frequency axis should begin at 0, and the axis should not be broken. All bars must have a common baseline to facilitate comparison. Categories of data should be presented in logical order and consistently with other figures and tables in the article. The baseline of a bar graph is not a coordinate axis and therefore should not have tick marks. Bar graphs may be used to compare frequencies between groups. In most cases, the number of bars in a grouped bar graph should not exceed 3. Colors or tones used to designate each group should be distinct. To ensure that bars in black-and-white figures are distinguishable, a contrast in shading of at least 30% for adjacent bars is suggested. Color or shades of gray should be used instead of patterns and crosshatching (eg, diagonal lines) on bars.

Pie Chart. Pie charts compare relationships among component parts. Categories are represented by sections, with the area of the section being proportional to the relative frequency of each category. Pie charts are used commonly in publications intended for lay audiences but should be avoided in scientific publications. The angular areas of the individual components of pie charts may be difficult to compare between pie charts. Usually, data depicted in pie charts can be summarized in the text or in a table.

Dot (Point) Graph. Dot or point graphs display quantitative data other than counts or frequencies on a single scaled axis according to categories on a baseline (the scaled axis may be horizontal or vertical). Like that in bar graphs, the baseline does not represent a scale and therefore does not contain tick marks.

Point estimates are represented by discrete data markers, preferably with error bars to designate variability (Figure 5). Dot or point graphs may be used to compare data between study groups, including positive and negative data values relative to a centrally located 0 baseline ("derivation graph"), paired data from single individuals, or pooled data in meta-analyses and other analyses that combine data from individual studies.

See pages 98 to 106 in the AMA Manual of Style 10th edition for additional information.

Figure 3 (adapted from Schneider et al (2010) Deutsches Arzteblatt International, 107 (44): 776-782)

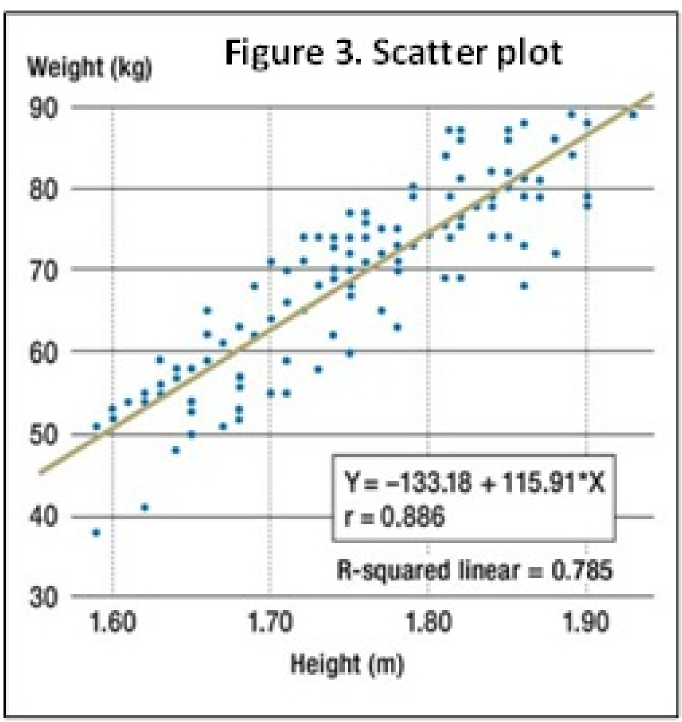


Figure 4 (adapted from Alexander et al (2004) JAMA, 292 (14):1696-1701)

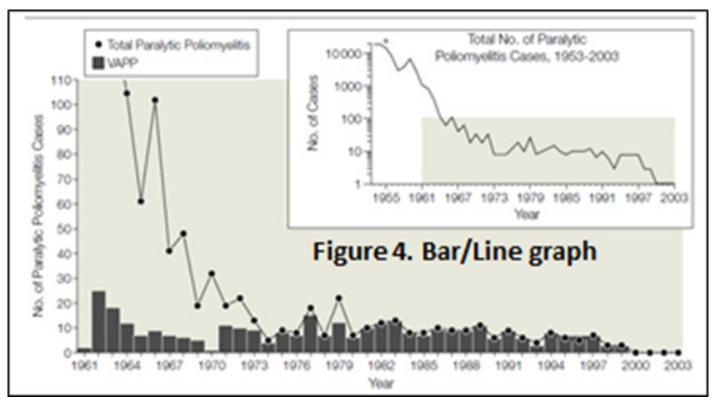
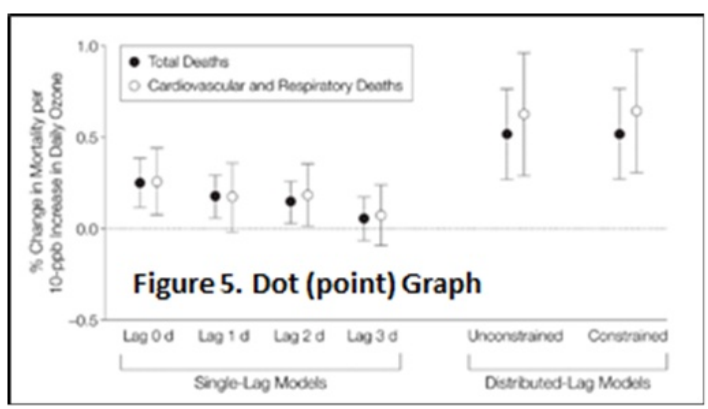


Figure 5 (adapted from Bell et al (2004) JAMA, 292(19):2372-2378)



How Do Grantwriters Measure Their Success

By Meg Bouvier, AMWA New England Chapter Member

I wrote my first grant application as a professional writer in about 2007. It was an NIH R01 resubmission, and I was lucky that it went from a score in the forty-odd percentile to 12%, and it was funded. Since that time, I have wrestled with the question of how to communicate to clients about my skill as a grantwriter (please note, I know writers occasionally post on the AMWA listservs stating that the word "grantwriter" is a misnomer. A grant is a monetary award and therefore cannot be written. Please forgive my expedient use of the word "grantwriter".)

I have employed numerous strategies over the years to try to quantify my grantwriting skills, some shamelessly pilfered from other writers and others I devised myself. In the first year or two, before I had many successful submissions, I had a sort of 'greatest hits' of my wins on my company's website, listing headings like "most improved score," "funded on first submission," and "least known funding mechanism."

A later iteration of my website boasted *recent* wins. I diligently listed R01s, R21s, and other mechanisms that I had helped clients land within the past 12-24 months. I am a strong believer that the NIH review process is a fairly rapidly moving target, therefore outdated grant experience is not terribly helpful.

By 2011, I had landed my first major center grant application. I was the lead writer on a $100 million HRSA grant application. It was the largest construction grant in university history, and my career was launched. This win was followed by a few more large-format wins on submissions for which I served as lead writer, so I simply listed the large center grants/contracts on my website and left it at that.

But I have begun to wonder what it all means. What should I count as a win? What if an R-series grantee only has me advise on a submission, or asks me for help only on the Aims, Significance, Innovation, and Intro? What if I only *edit* the submission? Do these count as *my* wins? I sometimes help a client write a first submission on which they might receive a percentile score in the high teens, then they land the grant on the resubmission, on which I did not work directly. Do I count *that* as my win, when the client credits me with the ultimate success of the submission? Nowadays, my marketing documents state that I have helped clients land over $200 million in federal funding. I figure I can justify that number based on the large-format wins for which I clearly wrote the applications. I can live with that number.

I have had clients ask me for my success rate—what percent of my clients land their grant? Seems like a benign and perfectly reasonable question, right? Nope. I would never suggest to a novice grantwriter that they maintain or advertise such a thing. Grantwriting is an iterative process. At NIH, few applications are funded on the first try, and it can take time to titrate a grantee's submission strategy. Success in one agency, IC, or study section does not mean you will be successful at others. It takes time, patience, legwork, and usually multiple submissions to figure it out. Most of my clients understand this and are willing to invest time and energy in developing their relationship with a given agency over time. And I thoroughly enjoy helping them in this process.

While one should write an application as if it were the only shot at funding, the grantwriter and client must also understand that a first submission to a new agency, study section, or IC will likely wind up being a learning experience. I find it very rewarding to work with a client over time as they develop their understanding of a given IC and study section, and build a relationship with a Program Officer. It is gratifying to help that client grow in terms of their NIH grantsmanship, and hopefully to land their grant on a subsequent submission and launch their relationship with NIH. The same holds true for experienced grantees looking to make the leap into center grants. It usually takes patience, hard work, a ridiculously thick skin, and multiple submissions to succeed.

No matter how strong the science, NIH statistics show that few funded projects are successful on the A0, i.e. first submission. Therefore, a grantwriter who boasts their funding rate is not likely to accept inexperienced applicants as clients—<u>yet these are the very clients who may benefit the most from our help</u>! If grantwriters only accept projects they know have a strong chance of funding, then who will help inexperienced grantees learn the ropes?

Perhaps the better measure of success in our field is if our clients feel we strengthened their application, educated them on the NIH grant process, and <u>improved their overall approach to grantsmanship</u>—skills they will carry with them throughout their career, whether they are successful on a given submission or not. The grantsmanship skills we teach them may also be passed on to the scientists our clients mentor, and to colleagues for whom they provide guidance on grant submissions.

I assume I do not need to state that you should

never guarantee success on a grant submission. Great writing and grantsmanship savvy are necessary, but not sufficient, to funding success at NIH. A grantwriter cannot change the science, and naturally many projects are not funded because of the science or because of a poor fit with the funding priorities of the granting agency. In addition, as grantwriters we cannot ensure that our clients will follow our advice, use our writing, or incorporate our edits.

If you are a grantwriter, how do you measure success? How do you capture that information and communicate it to potential clients? Most importantly, what is your goal as a grantwriter? Goodness knows, I am an extremely competitive person by nature, just ask my family. But like me, perhaps after thinking about it, you will find that going for the win is not necessarily your primary goal as a professional grantwriter.


MEG BOUVIER, PhD, is the owner of Meg Bouvier Medical Writing (www.megbouvier.com, a company that assists clients in writing persuasively about their biomedical research. She has helped clients land more than $200 million in federal funds, but is more proud that she has helped hundreds of clients improve their overall grantsmanship and feel less terrified of the NIH submission process. She was a press, policy, and communications writer for Dr. Francis Collins at the National Institutes of Health after completing an IRTA fellowship at NINDS. She holds a PhD in Biomedical Sciences from the Mt. Sinai School of Medicine.

A Disease by Another Name

By Rebecca J. Anderson, PhD, AMWA Pacific Southwest Chapter Member

The World Health Organization recently issued "best practice" guidelines for naming human infectious diseases. The new guidelines were prompted by the widespread habit of scientists, the media, and the general public to ascribe unofficial names to diseases. Those catchy nicknames are easy to pronounce and remember, and as such they are often "officially" adopted. But they may be inadvertently offensive, especially when describing diseases that emerge without warning, are not well understood, and may be life-threatening (think: Black Plague). According to WHO, "certain disease names provoke a backlash against members of particular religious or ethnic communities, create unjustified barriers to travel, and trigger needless slaughtering of food animals."

For example, swine flu isn't transmitted by pigs, but during a recent outbreak, some countries banned pork imports. Also, you may remember the push-back from veterans who were not at all happy that a certain infection (discovered in a Philadelphia hotel during their convention and sickening many of them) was labeled Legionnaires disease. The name stuck. Clearly as medical writers, we should be mindful of these unintended consequences.

WHO recommends that we avoid disease names that include a geographic location, people's names, species of animal or food, and references to specific cultures, populations, industries, and occupations. We should also avoid words that incite undue fear (e.g., death, fatal, epidemic). That doesn't leave much. In the words of one infectious disease expert, "The WHO document is laudable in its intent, but slightly daft."

Fortunately, WHO does offer some politically correct alternatives. It's ok to use words that characterize age group (pediatric, maternal), time course (acute, chronic, progressive, contagious), severity (severe, mild), and seasonality (summer, seasonal). Also, (thank goodness!) the new WHO guidelines are not retroactive and cover only infectious diseases. But I couldn't help wondering how we might have applied these rules to diseases that now have well-established nicknames.

Here are some examples (in alphabetical order), changed to comply with WHO's recommendations. (WARNING: these examples are only for illustrative purposes and, thankfully, not sanctioned by any organization)

Alzheimer's disease: progressive old people's syndrome (or Pops)
chickenpox: non-measles, non-rubella, non-smallpox, pox
Crohn's disease: chronic alimentary propulsive syndrome (or CRAPs)
Ebola (from the Ebola River in the Congo): severe contagious hemorrhagic fever
Elephantiasis: parasitic leg lymph hypertrophy
Hantavirus (from the Hantan River in South Korea): gripping airway syndrome-pulmonary (or Gasp)
Japanese encephalitis: summer flaviviral encephalitis
Kaposi's sarcoma: purple spots and bumps
Kawasaki disease: non-motorcycle idiopathic febrile syndrome
Lyme disease (from Old Lyme, Connecticut): tick vector radial red rash
mad cow disease: malicious zoonotic prion spongiform encephalopathy
miner's lung: chronic environmental pneumoconiosis
Montezuma's revenge: entero-toxigenic gastrointestinal distress
Rocky Mountain spotted fever: acute rickettsia rash and fever (or Arrf)
Sjögren's syndrome: tear-less, spit-less syndrome
Stevens-Johnson syndrome: idiopathic progressive exfoliative necrosis
Toxic shock syndrome: tampon fever
Tularemia (for Tulare County, CA): severe infectious ulceroglandular fever
West Nile virus: arthropod-borne encephalitis

The WHO guideline is available at: http://apps.who.int/iris/bitstream/10665/163636/1/WHO_HSE_FOS_15.1_eng.pdf?ua=1

REBECCA J ANDERSON, PhD, is a freelance medical writer and the author of two books, *Nevirapine and the Quest to End Pediatric AIDS* and *Career Opportunities in Clinical Drug Research*. Prior to medical writing, Dr. Anderson managed research and development projects for twenty-five years in the pharmaceutical/biotech industry. She holds a Ph.D. in pharmacology from Georgetown University. She lives in Southern California, and when she is not writing, she absorbs the sights and sounds of the West Coast's rich culture and heritage. She can be reached at rebeccanderson@msn.com.

Ask APRIL: What is Summer (Work) Style?

By April Reynolds, MS, ELS, AMWA Pacific Southwest Chapter Member

Career expert Lindsey Pollak says to dress for your profession.[1,2] But what does that mean for medical writers and editors interviewing or attending a conference? And what does that mean when it's hot outside?

Consider your audience

I polled my writers' group about topics for this month's column. Overwhelmingly, they wanted tips on how to dress professionally during the scorching summer months. I came up with some tips but wasn't sure how to give realistic and useful advice, considering that I haven't dressed up during summer for years. Then it happened: my computer died mid-July, and I had to go to my client's corporate site for repairs.

Connect with your audience

Here's what I wore:

a) White short-sleeve, button-up shirt with collar
b) Lightweight, dark-wash, trouser-cut jeans
c) Peep-toe (meaning a small opening, not a full open-toe) sling-back flats
d) Hair in a sleek ponytail

Here's why I wore it:

a) A crisp white shirt means business, but a short-sleeve white shirt is more appropriate for summer.
b) Nice jeans are fine on a Friday (which it was).
c) Whereas an open-toe shoe may be too casual, a peep-toe is dressier while still allowing your feet to breathe.
d) It was humid, and I have frizzy hair. This was my safest bet.

Create an outline

Before any of this took place, however, I used an invaluable tactic: planning. Because I knew my morning would be hectic—getting my child ready, getting myself ready, and getting the right amount of coffee to enable all of this to happen—I had a pretty solid idea of what I was going to wear ahead of time. I only needed to interact with IT, so I could be somewhat casual; however, you just never know who you'll run into. I stuck with the Girl Scouts on this one and was sure to "be prepared."

Planning can also be helpful when you're shopping for summer-friendly attire (or any attire, really). I try to have an idea of what I want before you go out to buy it. That way, I don't fall victim to markdowns or get overwhelmed. (My motto: If you wouldn't pay full price for it, you shouldn't buy it on sale.)

Enlist the help of a good editor

Start by looking on Pinterest. (I created a Pinterest board to give you some ideas: https://www.pinterest.com/writecorrect/summer-work-styles/) Put together an outfit in your head and then go out and try to recreate it as close as you can, and within your budget.

Admittedly, Pinterest tends to cater to young, thin, ultra-stylish women. But what about the rest of us? And what about men? Here re some additional suggestions:

MEN

DO	DON'T
• Light-weight pant in khaki or navy pants	• Black pants
• Jeans: dark blue wash or other neutral color, like gray or beige	• Jeans with holes, faded, ill-fitting (ie, "dad jeans")
• Light-weight cotton bottom-up shirt with conservative (but fun) pattern like Gingham	• Dark wool or heavy-weight dress or suit
• Short-sleeve button-up	• T-shirt
• Boat shoe or driving moccasin	• Flip flops, Tevas, or other "sporty" footwear
• Groom (no 5 o'clock shadow unless it's well kempt)	• Overdo it on cologne; lighter is better for summer (orange or lemon essence)
• Anti-sweat products like Body Glide, baby	• Let sweat get the best of you

WOMEN

DO	DON'T
• Conservative-print dress in lighter-weight fabric (cotton with stretch to combat wrinkles)	• Heavy black shoes or boots
• Sling-back kitten heel or wedge (not espadrille)	• Jeans (unless dark color worn on a Friday with a light blazer or other dressy piece)
• Short-sleeve blouse, brown pant, silk scarf around neck	• Dark wool or heavy-weight dress or suit
• Bare leg	• Or if you need coverage, consider a pop of color instead of black or bare
• Linen blazer or cotton cardigan to carry or wrap around shoulders, especially with a sleeveless dress/blouse	• Spaghetti-strap dress or dress that hits above the knee; keep shapes conservative (ie, shift dress)
• Body Glide anti-sweat stick—anywhere sweating may be a problem—or powder in your shoes to help combat sweat	• Flip flops, Tevas, or other "sporty" footwear
• Groom, ie, style your hair (bun, formal ponytail, etc)	• Makeup that will run or bleed, like too much mascara or eyebrow pencil; try waterproof versions instead
• Anti-sweat products like Body Glide, baby powder, and blotting papers	• Sweat through your clothes

Also, because I worked in retail during college, I always suggest engaging the help of salespeople. They work with the merchandise every day, and they can see you with an objective eye. Nordstrom has personal shoppers, and Levi's has a great fit guide for men.[3]

Make a strong statement

Sure, clothes aren't everything; it's the quality of our work that's important. But the idea is to dress like you're worth the money you're asking your client to pay—while never letting them see you sweat.

REFERENCES:

1. http://www.glamour.com/fashion/blogs/dressed/2013/06/outfit-idea-what-to-wear-on-a
2. http://www.lindseypollak.com/wear-this-not-that-a-millennials-guide-to-business-casual/
3. http://www.levis.com.au/men-fit-guide

APRIL REYNOLDS, MS, ELS, is a medical writer & editor and the president of Write/Correct, Inc. She has published works on topics that range from jeans (for fashion magazines) to genes (for medical publications). She lives in San Diego with her husband and son.

Medical writing's own fashion experimenter and amateur decorator answers your style questions.

Email your questions to: AskAprilatAMWA@gmail.com.

Chapter Happy Hour in San Diego, June 18, 2015

About 20 members met for drinks and networking at Bella Vista Social Club and Caffe in La Jolla last month. The chapter would like to thank Asoka Banno and Brea Midthune for initiating and organizing the event. Thank you also to Real Life Sciences, a staffing consultancy, and Recruiting Consultant Conor Trombetta for sponsoring appetizers and drinks at the event. In the pictures: Jennifer Veevers, ??, Conor Trombetta (top left); Julian Kaye, ???, Noelle Demas (top right); Brea Midthune, Andrew Hellman, Anna Larocca, Valerie Breda (bottom left); and ?? (bottom right).

AMWA Pacific Southwest chapter warmly welcomes our new members

Carol Koprowski – Encino

Chip Reuben – Redondo Beach

C. Corrigan – Rancho Santa Margarita

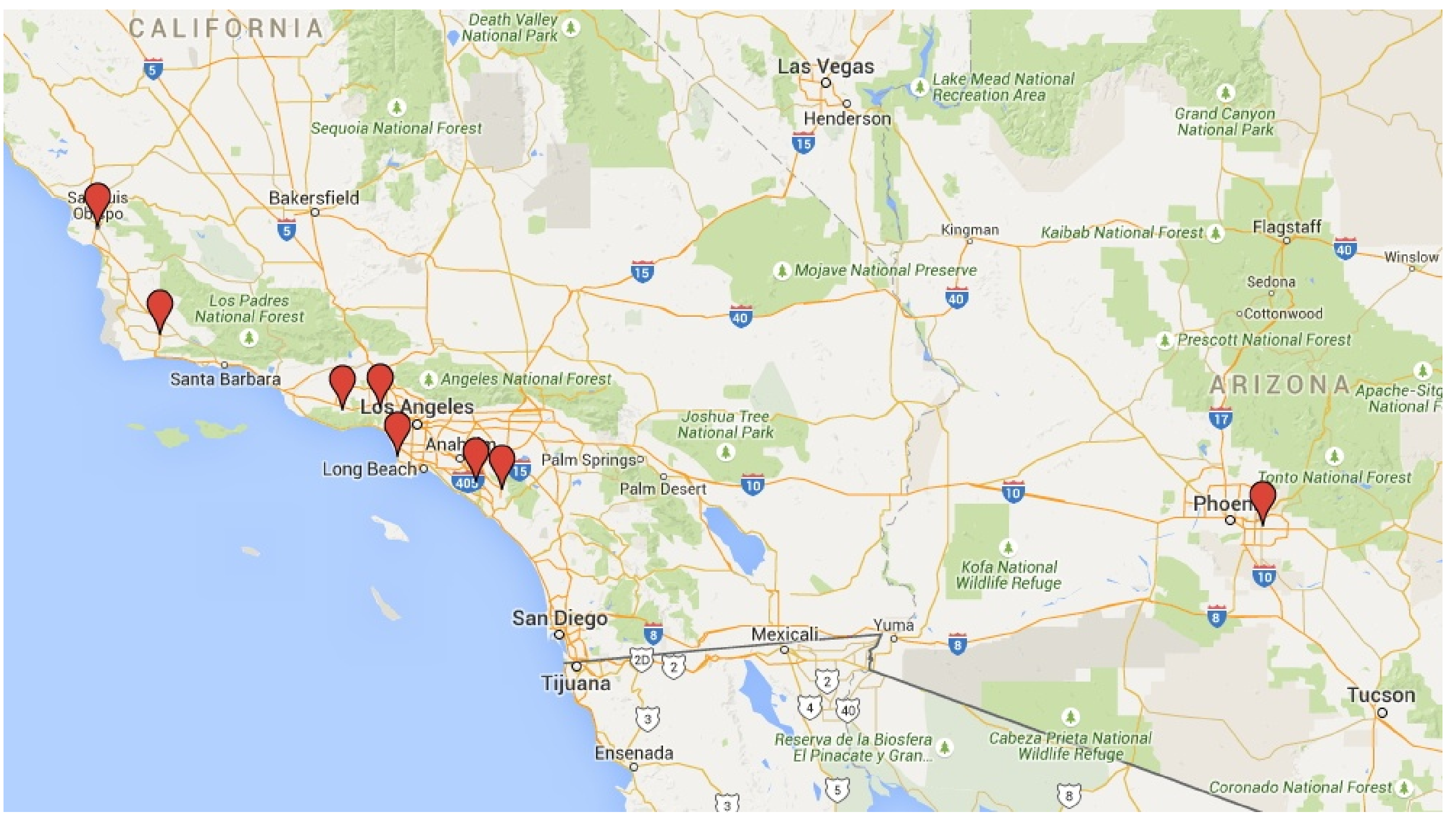
Dora Saltzman – San Luis Obispo

Michelle Zakson – Westlake Village

Rachel Meyer – Mesa, AZ

Robin Alexander – Irvine

Sharon Kelly – Buellton



List courtesy of Gail Flores, PhD, AMWA Pacific Southwest chapter membership coordinator.

Chapter Events' Calendar

August 15, 2015. Joint meeting with SDRAN to discuss Writing for a Patient Audience.

September 19, 2015 Chapter organized symposium: "Medical Writers' Toolbox Decoded". Location: Thousand Oaks, CA (Amgen) -- SAVE THE DATE

SDRAN SAN DIEGO REGULATORY AFFAIRS NETWORK

AMERICAN MEDICAL WRITERS ASSOCIATION

Saturday, August 15, 2015

Registration

Flyer

"Writing with the End User in Mind: Communication and Labeling to Improve Patient Safety"

Time:
11:30 am – 12:30 pm, Registration, Open Networking, Lunch
12:30 pm – 1:30 pm, Speaker: Joely Gardner
1:30 pm – 1:45 pm, break
1:45 pm – 2:45 pm, Speaker: Tim Vanderveen
2:45 pm – 3:30 pm, CareFusion Safety Center tour

Location:
CareFusion (directions on last page of this flyer)
3750 Torrey View Court
San Diego, CA 92121

Program Speakers:
Program Speaker: Joely Gardner
Program Speaker: Tim Vanderveen
Tour: CareFusion Center for Safety and Clinical Excellence

Program Summary:
The medical device industry is seeing many more recalls based on poor usability. This has led the FDA to require usability testing as part of design validation. The requirements for usability also apply to IFUs and labels. In fact, these requirements are global. Improving patient safety is at the heart of the regulators' focus on usability and improvements in product labeling for the intended user.

REGISTRATION: https://s08.123signup.com/servlet/SignUpMember?PG=1534811182300&P=15348111911429654800

FLYER:
http://media.wix.com/ugd/b7c3b3_a9d9fada904640168660d1d2780465f0.pdf?dn=August%2015%20SDRAN_AMWA%20Program%20Flye.pdf

AMERICAN MEDICAL WRITERS ASSOCIATION AMWA Pacific Southwest Chapter	**September 19, 2015** (Saturday) **Amgen, Inc.** **One Amgen Center Dr** **Building 24 Conference Center Auditorium** **Thousand Oaks, CA 91320**

AMWA Pacific Southwest Chapter presents:

Medical Writers' Toolbox Decoded

The skills and expertise needed for excelling as a medical writer are diverse and depend on the industry and job function. While there is no substitute for hands-on experience, this symposium is an opportunity to get a flavor of **how medical writers practice their craft**.

Join us for a day of interactive lectures, demonstrations, and discussions to learn about:

- Guidelines, styles guides, and templates used by medical writers in clinical writing and publication development
- Illustrations and data presentation tools and tricks
- Reference library tools and information management strategies in a pharmaceutical environment.

Symposium Program

10:00 AM – 10:30 AM	Coffee and Registration
10:30 AM – 11:30 AM	**Style Guides and Best Practices** ***Dikran Toroser, PhD, CMPP*** Medical Writing Senior Manager, Amgen, Inc AMA-zing Style column (creator and contributor), *Postscripts*
11:30 AM – 12:30 PM	**Developing Figures and Illustrations for Publications** ***Annalise M Nawrocki, PhD*** Medical Writing Manager, Amgen, Inc
12:30 AM – 1:30 PM	Lunch, Networking
1:30 PM – 3:00 PM	**The Right Information at the Right Time: The New Pharmaceutical Library** ***Christopher Mundy, PMP*** Knowledge Strategy Consultant, CM Consulting
3:00 PM– 3:45 PM/close	Q & A, Networking

About the Speakers:

DIKRAN TOROSER, PhD, CMPP, a member of the AMWA Pacific Southwest chapter, is a regular contributor to the Postscripts magazine since 2012. He developed the monthly AMAzing Style column which covers topics from the AMA Manual of Style, and has also written on publication-related topics in these pages. Dikran is currently a Senior Medical Writing Manager at Amgen Inc. in Thousand Oaks, California. He earned his PhD in Biochemistry from Newcastle University (UK), and did his post-doctoral training in biochemical genetics at the John Innes Center of the Cambridge Laboratory (Norwich, UK) and in molecular biology with the USDA. Prior to Amgen, Dikran was on the faculty (research) at the School of Pharmacy at the University of Southern California. He can be reached at dtoroser@amgen.com.

ANNALISE M NAWROCKI, PhD, is a member of the Pacific Southwest Chapter of the American Medical Writers Association. Annalise holds a BA in Molecular Biology and Genetics (with a minor in English Literature) from Northwestern University, and earned her PhD in Ecology and Evolutionary Biology from the University of Kansas. She is currently a Medical Writing Manager at Amgen Inc. in Thousand Oaks, California, where she develops publications and conference presentations in the cardiovascular therapeutic area. She can be reached at nawrocki@amgen.com.

CHRISTOPHER MUNDY, MS, PMP, is Principal and Knowledge Strategy Consultant at CM Consulting, San Francisco Bay Area. He works with start-ups, biotech and pharmaceutical companies to conduct information gap analyses, review the competitive intelligence landscape, and develop corporate library solutions. He has implemented integrated information systems managing corporate records, scientific, research, clinical and regulatory records linked to cloud systems providing on-demand access to functional groups within and outside the organization. He has several project management credentials under his belt including the Six Sigma Green Belt. He is pursuing his Master's Degree in Information and Knowledge Strategy from Columbia University, New York. He can be reached via LinkedIn: https://www.linkedin.com/in/christophermmundy

Registration: Register at https://www.123signup.com/event?id=pxfjd

Register by September 4, 2015
Registration limited to 150 participants

Cost: Free (courtesy of Amgen)
Lunch: Provided by Amgen

AMWA Chapter Contacts (regarding symposium):

Ajay Malik, PhD, ajay@amwa-pacsw.org
Donna Simcoe, MS Biotech, MS Med Writing, MBA, CMPP, president@amwa-pacsw.org

Amgen Organizing Committee:

Jenilyn J. Virrey, PhD, Medical Writing Senior Manager, jvirrey@amgen.com
Albert Rhee, PhD, Medical Writing Manager, arhee@amgen.com
Laura Bruce, Administrative Coordinator, laura.bruce@amgen.com

Medical Writing Open Positions

Compiled By: **Sharyn Batey, PharmD, MSPH**

Employment Coordinator, AMWA Pacific Southwest Chapter

Technical Writer/Editor
Undisclosed Company in Phoenix, AZ
Recruiter: Sterling-Hoffman Executive Search
http://www.mybiotechcareer.com/JD/Medical-Writing-Arizona-Biotechnology-Jobs-Careers-8659

Freelance Technical Writer
Dermalogica, Carson, CA
http://chc.tbe.taleo.net/chc04/ats/careers/requisition.jsp?org=DERMA&cws=2&rid=2081&source=Indeed

Medical Writer - Pharmaceutical
Brandkarma, Irvine, CA
http://job-openings.monster.com/monster/8af005f9-c97d-4303-b0e9-a0ce8b5b4e77?mescoid=2700440001001&jobPosition=5

Medical Writing Specialist
Covidien, Irvine, CA
http://job-openings.monster.com/monster/65f17c9b-d7df-4976-8186-87f8568e8034?mescoid=2700440001001&jobPosition=7

Medical Writer
Sonendo, Inc., Laguna Hills, CA
https://www.smartrecruiters.com/SonendoInc/82527360-medical-writer

Technical Writer, Biotech
Sequoia, Oceanside, CA
https://www.smartrecruiters.com/Sequoia/83365813-technical-writer-biotech

Principal Medical Writer
Ardea Biosciences, Inc., San Diego, CA
http://job-openings.monster.com/monster/55747035-25f3-4ae9-9758-0174ef402a1c?mescoid=2700440001001&jobPosition=14

Senior/Principal Medical Writer
Intercept Pharmaceuticals, San Diego, CA
http://interceptpharma.submit4jobs.com/index.cfm?fuseaction=85416.viewjobdetail&CID=85416&JID=196172&source=Indeed

Clinical Document Specialist
Neurocrine Biosciences, Inc., San Diego, CA
http://www.biospace.com/jobs/job-listing/clinical-document-specialist-346338

*Note: Occasionally weblinks in the PDF document may not work if the web address is long and splits into 2 lines. You may copy and paste the complete link into a new browser tab or window to reach the correct website.

Senior Medical Writer
Neurocrine Biosciences, Inc., San Diego, CA
http://www.biospace.com/jobs/job-listing/sr-medical-writer-346339

Senior Medical Writer
Nuvasive, San Diego, CA
http://job-openings.monster.com/monster/5f840a7a-a24b-4685-baaa-f9c46ef62d54?mescoid=2700440001001&jobPosition=17

Medical Writing Associate Director
Receptos, San Diego, CA
https://receptos.hyrell.com/UI/Views/Applicant/VirtualStepPositionDetails.aspx?TemplateId=162444&IsAutoRefresh=True&r=Indeed&tzi=Pacific%20Standard%20Time

Medical Writer
Undisclosed Company in South Region of California
Recruiter: Sterling-Hoffman Executive Search
http://www.mybiotechcareer.com/JD/Clinical-Research-Affairs-R-AND-D-Science-Medical-Affairs-California-South-Region-Biotechnology-Jobs-Careers-9911

Medical Writer / Diabetes - Outcomes Surveys
Undisclosed Company in San Diego, CA
Recruiter: Writing Assistance, Inc.
http://mh188.maxhire.net/cp/?E8556C361D43515B7E59126532571C69482D7C48&AspxAutoDetectCookieSupport=1

Senior Associate Regulatory Writing
Amgen, Thousand Oaks, CA
https://sjobs.brassring.com/TGWebHost/jobdetails.aspx?jobId=1145986&partnerid=25236&siteid=5308&codes=JB_Indeed

Regulatory Writing Manager
Amgen, Thousand Oaks, CA
http://job-openings.monster.com/monster/1611bcb9-cdc3-4878-9e73-33c7b19df8fa?mescoid=2700439001001&jobPosition=11

Regulatory Writing Senior Manager
Amgen, Thousand Oaks, CA
http://job-openings.monster.com/monster/ca774dcb-405a-40b2-b549-1ec0d5e5ef5a?mescoid=2700439001001&jobPosition=13

Medical Writer (Protocols and CSRs) - Remote
Lotus Clinical Research, LLC, Pasadena, CA
http://lotuscr.com

If you want to share job leads with the members of the Pacific Southwest Chapter, please contact Sharyn at employment-coordinator@amwa-pacsw.org.

Sir Ronald Fisher, Statistician, Geneticist, and a Farmer

Stained glass window in the dining hall of Caius College, Cambridge, commemorating Ronald Fisher and representing a Latin square. CREDIT: Wikipedia

Sir Ronald Aylmer Fisher, FRS, was a statistician and geneticist whose influence extends from population and evolutionary biology to psychology and agricultural research. He is the founder of modern discipline of statistics introducing numerous concepts, including analysis of variance (ANOVA), Fisher's z-distribution (F distribution), maximum likelihood estimations, and many more. He transformed the fields of psychology, agricultural research, and genetics by introducing new statistical methods and introducing the use of mathematical models.

His first book, "Statistical Methods for Research Workers," first published in 1925 remains an influential text in the field.[1] This book advanced the principles of "design of experiments"—later (in 1935), he published another book with the same name.[2] His work is reflected in the way we design and conduct clinical trials today using the foundations he created for designing of experiments with placebo group and null hypothesis, and using statistical analysis tools, such as, analysis of variance, statistical inference, etc. *The time and effort invested at the front end of a clinical trial is more important than data analysis after the study is over*. He once said: "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of." [3]

Fun Facts: Due to poor eyesight, he learned mathematics by being read aloud. After being told that he can't follow his career in mathematics, he lived as subsistence farmer for 2 years. Those 2 years were the launch pad to his contributions to the field of agricultural research, evolutionary biology, and applied statistics.

References

1. Fisher RA. Statistical methods for research workers (14th ed.). Oliver and Boyd, Edinburgh. (1970) [1925] ISBN-10: 0-05-002170-2, 0-02-844730-1
2. Fisher RA. The Design of Experiments (9th ed.). Macmillan. (1971) [1935]. ISBN-10: 0-02-844690-9
3. Presidential Address to the First Indian Statistical Congress (1938)

Sources and Further Reading

Wikipedia. https://en.wikipedia.org/wiki/Ronald_Fisher
Mejdal A. Celebrating statisticians: Ronald A. Fisher. JMP Blog. 2014 Jan. http://blogs.sas.com/content/jmp/2013/01/07/celebrating-statistics-statisticians-ronald-a-fisher/
Savage LJ. On Rereading R. A. Fisher. Ann Statist. 1976; 4(3):441-500. http://projecteuclid.org/euclid.aos/1176343456

—Editor